



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(3) International Patent Classification: 6:

C12N 15/12, 5/10, 1/21, C07K 14/47,  
16/18, C12N 1/21, C07K 14/47, 16/18,  
C12Q 1/68, G01N 33/50, 33/53, 33/68,  
A61K 38/17

A2

(11) International Publication Number: WO 98/39448  
(43) International Publication Date: 11 September 1998 (11.09.98)

(21) International Application Number: PCT/US98/04493

(22) International Filing Date: 6 March 1998 (06.03.98)

(30) Priority Date:

60040,162 7 March 1997 (07.03.97) US  
60040,333 7 March 1997 (07.03.97) US  
60038,621 7 March 1997 (07.03.97) US  
60040,161 7 March 1997 (07.03.97) US  
60040,626 7 March 1997 (07.03.97) US  
60040,336 7 March 1997 (07.03.97) US  
60040,163 7 March 1997 (07.03.97) US  
60043,580 11 April 1997 (11.04.97) US  
60043,568 11 April 1997 (11.04.97) US

(Continued on the following page)

(71) Applicant (for all designated States except US): HUMAN  
GENOME SCIENCES, INC. (US/US); 9410 Key West  
Avenue, Rockville, MD 20850 (US).

(72) Inventors, and

(US/US); 18528 Heritage Hills Drive, Olney, MD 20832  
(US) ROSEN, Craig, A. (US/US); 22400 Kolling Hills  
Road, Laytonville, MD 20882 (US); FISCHER, Carrie, L.  
(US/US); 5810 Hall Street, Burke, VA 22015 (US); SOP-  
PERT, Daniel, R. (US/US); 15050 Shilfield, Place, Cente-  
ville, VA 22020 (US); CARTER, Kenneth, C. (US/US);  
11601 Brandy Hall Lane, North Potomac, MD 20878  
(US); BEDNARIK, Daniel, P. (US/US); 8822 Blue Sea  
Drive, Columbia, MD 21046 (US); ENDRESS, Gregory, A.  
(US/US); 9729 Clagett Farm Drive, Pocomoke, MD 20854  
(US); YU, Guo-Liang (CN/US); 13524 Shaw Blue Lane,  
Darnestown, MD 20878 (US); NI, Jian (CN/US); 5502  
Manorfield Road, Rockville, MD 20853 (US); FENG, Ping  
(CN/US); 4 Reida Court, Gaithersburg, MD 20878 (US).  
YOUNG, Paul, E. (US/US); 122 Beckwith Street, Gaithers-  
burg, MD 20878 (US); GREENE, John, M. (US/US); 872  
Diamond Drive, Gaithersburg, MD 20878 (US); FERRIER,  
Ann, M. (US/US); 13203 L Astoria Hill Court, Germantown,(74) Agents: BROOKES, Andrew, A. et al.; Human Genome  
Sciences, Inc., 9410 Key West Avenue, Rockville, MD  
20850 (US).(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR,  
BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE,  
GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ,  
LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW,  
MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TM, TR, TT, UA, UG, UZ, VN, YU, YZ, ZW, AR, AO,  
AT, BE, BG, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT,  
LU, MC, NL, PT, SE, SD, SI, SK, SL, SM, SN, TD, TO).  
Published  
Without international search report and to be republished  
upon receipt of that report.(54) Title: 186 HUMAN SECRETED PROTEINS  
(57) Abstract

The present invention relates to 186 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

60043,314	11 April 1997 (11.04.97)	US	60047,596	23 May 1997 (23.05.97)	US	60056,637	22 August 1997 (22.08.97)	US
60043,569	11 April 1997 (11.04.97)	US	60047,612	23 May 1997 (23.05.97)	US	60056,903	22 August 1997 (22.08.97)	US
60043,311	11 April 1997 (11.04.97)	US	60047,632	23 May 1997 (23.05.97)	US	60056,888	22 August 1997 (22.08.97)	US
60043,674	11 April 1997 (11.04.97)	US	60047,601	23 May 1997 (23.05.97)	US	60056,879	22 August 1997 (22.08.97)	US
60043,669	11 April 1997 (11.04.97)	US	60047,595	23 May 1997 (23.05.97)	US	60056,880	22 August 1997 (22.08.97)	US
60043,312	11 April 1997 (11.04.97)	US	60047,588	23 May 1997 (23.05.97)	US	60056,894	22 August 1997 (22.08.97)	US
60043,313	11 April 1997 (11.04.97)	US	60047,585	23 May 1997 (23.05.97)	US	60056,911	22 August 1997 (22.08.97)	US
60043,672	11 April 1997 (11.04.97)	US	60047,586	23 May 1997 (23.05.97)	US	60056,874	22 August 1997 (22.08.97)	US
60043,576	11 April 1997 (11.04.97)	US	60047,590	23 May 1997 (23.05.97)	US	60056,910	22 August 1997 (22.08.97)	US
60043,670	11 April 1997 (11.04.97)	US	60047,589	23 May 1997 (23.05.97)	US	60056,882	22 August 1997 (22.08.97)	US
60047,600	23 May 1997 (23.05.97)	US	60047,593	23 May 1997 (23.05.97)	US	60056,631	22 August 1997 (22.08.97)	US
60047,615	23 May 1997 (23.05.97)	US	60047,614	23 May 1997 (23.05.97)	US	60056,945	22 August 1997 (22.08.97)	US
60047,597	23 May 1997 (23.05.97)	US	60047,591	23 May 1997 (23.05.97)	US	60056,632	22 August 1997 (22.08.97)	US
60047,502	23 May 1997 (23.05.97)	US	60048,974	06 June 1997 (06.06.97)	US	60056,654	22 August 1997 (22.08.97)	US
60047,583	23 May 1997 (23.05.97)	US	60048,994	06 June 1997 (06.06.97)	US	60056,876	22 August 1997 (22.08.97)	US
60047,582	23 May 1997 (23.05.97)	US	60049,610	13 June 1997 (13.06.97)	US	60056,881	22 August 1997 (22.08.97)	US
60047,617	23 May 1997 (23.05.97)	US	60051,926	16 July 1997 (16.07.97)	US	60056,909	22 August 1997 (22.08.97)	US
60047,584	23 May 1997 (23.05.97)	US	60052,874	16 July 1997 (16.07.97)	US	60056,875	22 August 1997 (22.08.97)	US
60047,581	23 May 1997 (23.05.97)	US	60053,724	18 August 1997 (18.08.97)	US	60056,862	22 August 1997 (22.08.97)	US
60047,592	23 May 1997 (23.05.97)	US	60056,886	22 August 1997 (22.08.97)	US	60056,887	22 August 1997 (22.08.97)	US
60047,587	23 May 1997 (23.05.97)	US	60056,887	22 August 1997 (22.08.97)	US	60056,908	22 August 1997 (22.08.97)	US
60047,492	23 May 1997 (23.05.97)	US	60056,890	22 August 1997 (22.08.97)	US	60056,884	22 August 1997 (22.08.97)	US
60047,598	23 May 1997 (23.05.97)	US	60056,878	22 August 1997 (22.08.97)	US	60057,669	05 September 1997 (05.09.97)	US
60047,585	23 May 1997 (23.05.97)	US	60056,862	22 August 1997 (22.08.97)	US	60058,785	12 September 1997 (12.09.97)	US
60047,582	23 May 1997 (23.05.97)	US	60056,872	22 August 1997 (22.08.97)	US	60061,060	02 October 1997 (02.10.97)	US
			60056,882	22 August 1997 (22.08.97)	US			

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Togo
BA	Bosnia and Herzegovina	GR	Greece	MD	Republic of Moldova	TM	Tajikistan
BB	Barbados	HN	Honduras	MG	Madagascar	TR	Turkey
BE	Belgium	ON	Ontario	ML	The former Yugoslav Republic of Macedonia	TT	Trinidad and Tobago
BG	Bulgaria	GR	Greece	MR	Mali	UA	Ukraine
BO	Bolivia	HU	Hungary	MT	Malta	UG	Uganda
BR	Brazil	IE	Ireland	MN	Mongolia	US	United States of America
BY	Belarus	IL	Israel	MR	Mauritania	UZ	Uzbekistan
CA	Canada	IS	Iceland	MY	Malaysia	VN	Viet Nam
CC	Cocos (Keeling) Islands	JP	Japan	NB	Nicaragua	ZW	Zimbabwe
CH	Switzerland	KG	Kyrgyzstan	NL	Netherlands		
CI	Cote d'Ivoire	KO	Korea	NZ	New Zealand		
CM	Cameroun	KP	Democratic People's Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russia Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## 186 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig



analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 12301 Park Lawn Drive, Rockville, Maryland 20852, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA contained within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to poly(A+ sequences (such as any 3' terminal poly(A+ tract of a cDNA shown in the sequence listing) or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polynucleotide or polynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, uridyated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods.

Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotyrosine/inositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, **PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES**, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); **POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS**, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., *Meth Enzymol* 182:626-646 (1990); Rattan et al., *Ann NY Acad Sci* 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table I.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## 25 Polynucleotides and Polypeptides of the Invention

### FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in testes tumor and to a lesser extent in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly of the testes, and defects of the central nervous system such as seizure and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly cancer of the testes and central nervous system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, brain and other tissue of the nervous system, and blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of testicular cancer and treatment of central nervous system disorders since this gene is primarily expressed in the testes tumor and developing brain.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in cancer tissues, such as breast cancer and Wilm's tumor, and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and/or tumors, particularly, those found in the breast, and developmental abnormalities or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and fetal tissue and, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 314 as residues: Pro-11 to Thr-18, Leu-43 to Pro-50, Gly-64 to Leu-72, and Leu-81 to Lys-86.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of cancers and/or tumors, particularly, those found in the breast since expression is mainly in cancer/tumor tissues. May serve as therapeutic proteins for proliferation/differentiation of fetal tissues.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 3**

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in spleen, chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or leukemias, diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders or leukemias, diseases of the immune system since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 4**

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or lymphocytic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 5**

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 317 as residues: Pro-13 to Lys-21.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 6**

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 318 as residues: Lys-31 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood diseases since it is expressed in tissues related to immune function.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 7

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system and brain associated diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders, immune diseases or brain associated diseases (specifically of the pineal gland) since expression is in tissues related to immune function.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 8

The translation product of this gene shares sequence homology with an organic cation transporter which is thought to be important in organic cation uptake in the kidney and liver. (See Accession No. 2343059.) Preferred polypeptide fragments comprise the amino acid sequence ITTAIQMCLVNXELYPTFVRNXGVMVCSLCDIGGIPT FIVRLREVWQALPULFAVLGALLAAGVTLPLPETKGVLPETMKDAENLGRKAKPKENTLYLK VQTSEPSGT (SEQ ID NO: 615) or TMKDAENLGRKAKPKENT (SEQ ID NO: 616) as well as N-terminal and C-terminal deletions of these fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic and renal diseases where drug elimination/cation exchange (organic cation uptake) in the liver and kidney are problematic. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 320 as residues: Asn-64 to Asn-74, and Gln-81 to Gly-87.

The tissue distribution and homology to organic cation transporter indicate that polynucleotides and polypeptides corresponding to this gene are useful as a polyspecific transporter that is important for drug elimination in the liver (and possibly kidney) since expression is found in the liver.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in eosinophil induced with IL-5 and to a lesser extent in fetal liver and spleen. This gene also maps to chromosome 15, and therefore can be used in linkage analysis as a marker for chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system, particularly allergies or asthma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosis of diseases involving eosinophil reactions since expression seems to be concentrated in eosinophils and other tissues involved in immunity such as the liver and spleen.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 10

This gene is expressed primarily in tissues of hematopoietic lineage and to a lesser extent in Hodgkins lymphoma. Any frame shifts in this sequence can easily be clarified using known molecular biology techniques.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and immune deficiency or dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, lymphoid and reticuloendothelial tissues, and cancerous tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/ diagnosis for lymphomas or immune dysfunction or as a therapeutic protein useful in immune modulation based on expression in anergic T-cells and lymphomas.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in neutrophils and to a lesser extent in activated lymphoid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the cell type present in a biological sample and for diagnosis of diseases and conditions: inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 323 as residues: Glu-40 to Lys-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for modulation of an immune reaction or as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in brain and to a lesser extent in activated T-cells. It is likely that the open reading frame containing the predicted signal peptide continues in the 5' direction. Preferred polypeptide fragments comprise the amino acid sequence PRVRNSPEDLGLSLTGDSCKL (SEQ ID NO:617).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disorders including ischemic shock, alzheimers and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain, and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 324 as residues: Ser-5 to Glu-14, Ile-21 to Pro-35, Ser-65 to Asp-81, Cys-89 to Val-96, Lys-136 to Ser-145, Ile-152 to Met-169, and Arg-189 to Lys-196.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic/treatment for cancers of the given tissue or in the treatment of neurological disorders of the CNS.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 13**

This gene was also recently cloned by other groups, naming this calcium-activated potassium channel gene, hKC4. (See Accession No. AF033021, see also, Accession No. 2584866.) This gene is mapped to human chromosome 19q13.2. A second signal sequence likely exists upstream from the predicted signal sequence as described in

Table 1. Preferred polypeptide fragments comprise: QADDLQATVAALCVLRGGGPWAG  
 SWLSPKTFGAMGGDLVLGLGALRRRKRL (SEQ NO: 618); or EQEKSLAGWALVLAXXGIGL  
 MVLHAEMLVFGGCSAVNATGHLSDTLVLPIPTFLTIGYGDVVPCTMWGKIVCLCTGVMGVCC  
 TALLVAVYARKLEFNKAEKHVHFMMDIQYTKEMKESAAARVLQEAWMFYKHTRRKESHAAR  
 XHQRXLLAAINAFRQVRLKHKRLREQVNSMVDISKMHMILYDLQQLSSSHRALEKQIDTLAG  
 KLDALTELLSTALGPRQLPEPSQSK (SEQ ID NO: 619), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in breast lymph node and T-cells, and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematologic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, blood cells and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 325 as residues: Arg-13 to Lys-23.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment/diagnosis of hematologic and diseases involving immune modulation based on distribution in the lymph node and T-cells.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 14**

This gene was recently cloned by another group, calling it PAPS synthase.

(See Accession No. e1204135.) Preferred polypeptide fragments comprise the amino acid sequence YQAHVSRNKRQGVVGTGRGFRGCTVWLTOLSGAGK (SEQ ID NO: 620).

Also preferred are the polynucleotide fragments encoding this polypeptide fragment. It has been discovered that this gene is expressed primarily in benign prostate hyperplasia, Human Umbilical Vein Endothelial Cells and to a lesser extent in smooth muscle and Human endometrial stromal cells-treated with estradiol.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammation, ischemia, and restenosis, based on endothelial cell and smooth muscle cell expression, and prostate diseases such as benign prostate hyperplasia or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate or vessels of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, endothelial cells, smooth muscle, and endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 326 as residues: Arg-21 to Asp-26, Lys-35 to Lys-44, Glu-49 to Asn-58.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosing diseases or conditions where the endothelial cell lining of the veins and arteries of underlying smooth muscle are involved.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 15**

This gene is expressed primarily in human 6 week embryo and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental anomalies or fetal deficiencies. Similarly, polypeptides and antibodies directed to these



polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly developmental in nature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 327 as residues Lys-50 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of developmental abnormalities.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 16

15 This gene is expressed primarily in kidney and amygdala and to a lesser extent in fetal tissues. This gene is mapped to chromosome 14, and therefore is useful in linkage analysis as a marker for chromosome 14.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) present in a biological sample and for diagnosis of diseases and conditions: kidney diseases, neurological disorders and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s). For a number of disorders of the above tissues, particularly of the renal system or developing fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, amygdala, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of conditions affecting the brain, kidneys and fetal development.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 17

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: solid tumors similar to ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 329 as residues Ser-51 to Val-56.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of solid tumors of the reproductive system such as ovarian cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in brain medulloblastoma. Preferred polypeptide fragments comprise the amino acid sequence: IRHEQHPNFSLEMSKSSLLFLPQLILLPVCAHLHELNC (SEQ ID NO: 643) and SPTSEKGHLLQAEHRHWVAGALVGVSGGLTLTTCSGPTKRPATKNYFLKRLQEMIRKAN (SEQ ID NO: 644), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the CNS or. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating medulloblastoma or similar tumors.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed primarily in adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose tissues expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating obesity by regulating the function and number of adipocytes

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in B cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of the immune system with an emphasis on B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumors of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

35

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of B cell derived tumors based on its expression in b cell lymphomas

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in immune cells and to a lesser extent in fetal tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cells of the immune system, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:333 as residues Asp-10 to Pro-19, Ser-74 to Tyr-79, Glu-95 to Lys-110.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of diseases involving alterations in T cell activity.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 22

It has been discovered that this gene is expressed primarily in ovarian tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumors of the reproductive organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian

30

35

and other reproductive tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 334 as residues: Leu-22 to Gln-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovarian tumors as it has only been identified in ovarian tumors.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 23

It has been discovered that this gene is expressed primarily in fetal tissues and to a lesser extent in osteoclastoma cell line

Therefore, polynucleotides and polypeptides of the invention are useful as

15

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteoporosis or arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment of conditions of abnormal bone remodeling due to enhanced activity of osteoclasts. This may be useful as a specific marker for malignancies derived from osteoclasts or their precursors.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with a periplasmic ribonuclease which is thought to be important in degrading extracellular polynucleotides

35

It has been discovered that this gene is expressed primarily in serum treated smooth muscle cells

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular disease such as restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Gln-30 to Lys-36, and Pro-41 to Arg-48.

15

The tissue distribution and homology to ribonucleases indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of pathological conditions of smooth muscle associated with bacterial or viral infiltration

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

20

This gene is expressed primarily in Early Stage Human Brain.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain development and related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

25

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain development and related diseases,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30

The tissue distribution and homology to this gene indicate that polynucleotides

and polypeptides corresponding to this gene are useful for diagnosis and treatment of

35

diseases affecting human brain development and related diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

It has been discovered that this gene is expressed primarily in human brain tissue.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases and other diseases related to brain diseases, which may be caused by brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain diseases and other diseases related.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 27**

It has been discovered that this gene is expressed primarily in Anergic T-cells.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases, inflammatory diseases and diseases related to T lymph cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

30 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune diseases, inflammatory diseases and diseases related to T lymph cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g.,

35 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for immune diseases, inflammatory diseases and diseases related to T lymph cells.

5

**FEATURES OF PROTEIN ENCODED BY GENE NO: 28**

The translation product of this gene shares sequence homology with *Shigella flexneri* positive transcriptional regulator CriR (criR) gene which is thought to be important in regulation of gene expression.

10

This gene is expressed primarily in human synovial sarcoma and normal human brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases particularly sarcomas of the synovium. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above

15 tissues or cells, particularly of the human brain and synovium and other related human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., synovial tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

20 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human synovial sarcoma and other related human brain diseases.

25

**FEATURES OF PROTEIN ENCODED BY GENE NO: 29**

This gene is expressed in bone marrow, infant brain, fetal liver and spleen, prostate and to a lesser extent in pineal gland, adipose tissue, kidney, adrenal gland, umbilical vein endothelial cells, and T cells.

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases related to bone marrow or

35

23

hematopoietic tissues, prostate, kidney, adrenal gland, and cardiovascular tissue or organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases related to hematopoietic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, hematopoietic cells, pineal gland, adipose tissue, kidney, adrenal gland, endothelial cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases related to hematopoietic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 30

This gene is expressed primarily in meningea and to a lesser extent in breast and adult brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases of the meningea and related brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the meningea and related brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., meningea, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

24

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the meningea and related brain diseases.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed in meningea, fetal spleen, osteoblast and to a lesser extent in activated T-cells, endometrial stromal cells, fetal lung, HL-60, thymus, testis and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: meningeal disease, osteoporosis, immune diseases, and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the meningeal diseases, osteoporosis, immune diseases, and hematopoietic diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endometrium, lung, thymus, testis, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of meningeal, osteoporosis, immune diseases, hematopoietic diseases, testis diseases and lung diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in human thymus and to a much lesser extent in infant brain, T-cells, smooth muscle, endothelial cells, bone marrow, human ovarian tumor and keratinocytes, testes, osteoclastoma, breast, and tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases involving the thymus, particularly thymic cancer and diseases involving T-cell maturation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the thymus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, brain, and other tissue of the nervous system, blood cells, bone marrow, ovaries, and testes, and other reproductive tissue, mammary tissue, tonsils, melanocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thymus particularly thymic cancer and diseases involving T-cell maturation.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human tonsils, and placenta, and to a lesser extent in adipocytes, melanocyte, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases, immune diseases, and obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory diseases, immune diseases, and obesity, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tonsils, placenta, adipocytes, melanocytes, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases such as inflammation, immune diseases, and obesity.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene is expressed in activated T cells, and to a lesser extent in pituitary, testis, and breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases relating to T cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary, testes and other reproductive tissue, mammary tissue, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of immune disorders.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 35

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain, and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 36**

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in human ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the ovarian disorders such as those involving germ cells, ovarian follicles, stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovariohathy.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 38**

This gene is expressed primarily in lymph node breast cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for used as a diagnostic marker for breast cancer.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 39**

This gene is expressed primarily in brain and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal disorders such as trauma, brain degeneration, and brain tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of neuronal disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 40

This gene is expressed in early stage human embryo, adrenal gland tumor, and immune tissues such as fetal liver, fetal spleen, T-cell, and myeloid progenitor cell line and to a lesser extent in ovary, colon cancer, and a few other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis including adrenal gland tumor, colon cancer and various other tumors, developmental and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, early stage human tissues, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, blood cells, bone marrow, ovary and other reproductive tissue, and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of immune and developmental disorders, and tumorigenesis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene is expressed primarily in fetal lung, endothelial cells, liver, thymus and a few other immune tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders such as immune deficiency and autoimmune diseases, pulmonary diseases, liver diseases, and tumor metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal lung, liver, endothelial cells, and immune tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues and cell types (e.g., lung, endothelial cells, liver, thymus, and other tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune disorders and pulmonary and hepatic diseases. Its promoter may also be used for immune system and lung-specific gene therapies. The expression of this gene in endothelial cells indicates that it may also involve in angiogenesis which therefore may play role in tumor metastasis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in liver, thyroid, parathyroid and to a lesser extent in fetal lung, stomach and early embryos.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic regulation, obesity, hepatic failure, hepatocellular tumors or thyroiditis and thyroid tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive/endocrine system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, thyroid, parathyroid, lung, stomach, and embryonic tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and the extracellular locations indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of digestive/endocrine disorders, including metabolic regulation, hepatic failure, malabsorption, gastritis and neoplasms.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 43**

This gene is expressed primarily in Schizophrenic adult brain, pituitary, front cortex, hypothalamus and to a lesser extent in retina, adipose and stomach cancer and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal tissue, adipose, stomach, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nerve system, including schizophrenia, neurodegeneration, and neoplasia. Additionally, a secreted protein in brain may serve as an endocrine.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 44**

The translation product of this gene shares sequence homology with GTP binding proteins which are thought to be important in signal transduction and protein transport.

This gene is expressed primarily in umbilical vein and microvascular endothelial cells, GM-CSF treated macrophage, anergic T cells, osteoblast, osteoclast, CD34+ cells and to a lesser extent in gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: bone formation and growth, osteonecrosis, osteoporosis, angiogenesis and/or hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoiesis systems, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues and cell types (e.g., endothelial cells, blood cells, bone, and gall bladder, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of bone formation and growth, osteonecrosis, osteoporosis, and/or hematopoiesis because its involvement in the growth signaling or angiogenesis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 45**

The translation product of this gene shares sequence homology with signal sequence receptor gamma subunit which is thought to be important in protein translocation on endoplasmic reticulum.

This gene is expressed primarily in adrenal gland, salivary gland, prostate, and to a lesser extent in endothelial cells and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: protein secretion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the secretory organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, salivary gland, prostate, endothelial cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to SSR gamma subunit indicate that polynucleotides and polypeptides corresponding to this gene are useful for endocrine disorders, prostate cancer, xerostomia or sialorrhea.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

This gene is expressed primarily in osteoclastoma cells and to a lesser extent in melanocyte, amygdala, brain, and stomach.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ossification, osteoporosis, fracture, osteonecrosis, osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, amygdala, brain and other tissue of the nervous system, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in intervention of ossification, osteoporosis, fracture, osteonecrosis and osteosarcoma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with proline rich proteins which is thought to be important in protein-protein interaction.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological and psychological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful in intervention

and detection of neurological diseases, including trauma, neoplasia, degenerative or metabolic conditions in the central nerve system. Additionally, the gene product may be secreted by the brain as an endocrine.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with the AOCB gene from *Aspergillus nidulans* which is important in asexual development.

This gene is expressed primarily in infant brain and to a lesser extent in the developing embryo, trachea tumors, B-cell lymphoma and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative diseases, leukemia and sarcoma's. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, blood cells, trachea, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in infant brain and sarcoma's and homology to a gene involved in a key step of eukaryotic development (fungal spore formation) indicates that the protein product of this clone could play a role in neurological diseases such as schizophrenia, particularly in infants. The existence of the gene in a B-cell lymphoma indicates the gene may be used in the treatment and detection of leukemia.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary disorders including lung cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in fetal lung indicates that it plays a key role in development of the pulmonary system. This would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung. It may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in hematopoietic cell types and fetal cells and to a lesser extent in all tissue types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in the immune system and hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene predominantly in hematopoietic cells and in the developing embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of lymphomas and disease states affecting the immune system or hematopoiesis disorders such as leukemia, AIDS, arthritis and asthma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene is expressed primarily in prostate and to a lesser extent in fetal spleen, fetal liver, infant brain and T cell leukemias.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate disorders, prostate cancer, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, and/or prostate gland expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, spleen, liver, brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in prostate indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection or treatment of prostate disorders or prostate cancer. Its distribution in fetal liver and fetal spleen indicates it may play a role in the immune system and its misregulation could lead to immune disorders such as leukemia, arthritis and asthma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 53

The translation product of this gene shares sequence homology with dynein.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuro-degenerative diseases of the brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly neuro-degenerative diseases expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution in the brain and homology to dynein, a microtubule motor protein involved in the positioning of cellular organelles and molecules indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection/treatment of neurodegenerative diseases, such as Alzheimers, Huntingtons, Parkinsons diseases and shizophrenia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with ubiquitin-conjugation protein, an enzyme which is thought to be important in the processing of the Huntingtons Disease causing gene.

This gene is expressed primarily in brain and to a lesser extent in activated macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disease states including Huntington's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of brain tissues. For a number of disorders of the above tissues or cells, particularly of the neurological systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution of this gene in the brain and its homology to a Huntington interacting protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the regulation of the expression of the Huntington disease gene and other neurodegenerative diseases including spinocerebellar ataxia types I and III, dentatorubropallidolysian and spinal bulbar muscular atrophy. In addition, the existence of elevated levels of free ubiquitin pools in Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis indicates that the ubiquitin pathway of protein degradation plays a role in these disease states. Thus, considering the gene described here is homologous to a ubiquitin-conjugation protein it may play a general role in neurodegenerative conditions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 56

This gene is expressed primarily in T-cells (anergic T-cells, resting T-Cells, apoptotic T-cells) and lymph node (breast), as well as brain (hypothalamus, hippocampus, pituitary, infant brain, early-stage brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune (e.g. immunodeficiencies, autoimmunities, inflammation, leukemias & lymphomas) and neurological (e.g. Alzheimer's disease, dementia, schizophrenia) disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous, hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, lymphoid tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention or detection of pathologies associated with the hematopoietic and immune systems, such as anemias (leukemias). In addition, the expression in brain (including fetal) might suggest a role in developmental brain defects, neuro-degenerative diseases or behavioral abnormalities (e.g. schizophrenia, Alzheimer's, dementia, depression, etc.).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in lung, and to a lesser extent in a variety of other hematological cell types (e.g. Raji cells, bone marrow cell line, activated monocytes).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary and/or hematological dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculo-pulmonary and hematopoietic systems, expression of this



39

gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful in the intervention and detection of pathologies associated with the vasculo-pulmonary system. In addition the expression of this gene in a variety of leukocytic cell types and a bone marrow cell line might suggest a role in hematopoietic and immune system disorders, such as leukemias & lymphomas, inflammation, immunodeficiencies and autoimmunities.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 58

15 The translation product of this gene shares sequence homology with adenylyate kinase isozyme 3 (gil163528 GTP:AMP phosphotransferase (EC 2.7.4.10) [Bos tauros]), which is thought to be important in catalyzing the phosphorylation of AMP to ADP in the presence of ATP or inorganic triphosphate.

20 This gene is expressed primarily in fetal liver, heart and placenta, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic, cardiovascular or reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic, cardiovascular and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, heart, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

35 corresponding to this gene are useful for the treatment and diagnosis of conditions related to hepatic function and pathogenesis, in particular, those dealing with liver development and the differentiation of hepatocyte progenitor cells.

40

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in CD34 positive cells (Cord Blood).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic differentiation and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the detection and treatment of conditions associated with CD34-positive cells, and therefore as a marker for cell differentiation in hematopoiesis, as well as immunological disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 60

The translation product of the predicted open reading frame of this contig has sequence identity to the murine gene designated Insulin-Like Growth Factor-Binding Protein (IGFBP)-1 as described by Lee and colleagues (Hepatology 19 (3), 656-665 (1994)).

This gene is expressed exclusively in hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hemangiopericytoma and other pericyte or endothelial cell proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and immune systems, expression of this gene at significantly higher or lower levels may routinely be detected in certain tissues and cell types (e.g., pericyte or endothelial cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Polynucleotides and polypeptides corresponding to this gene are useful as cell growth regulators since IGFBP-1-like molecules function as modulators of insulin-like growth factor activity. In addition, since IGFBP-1 is expressed at high levels following hepatectomy and during fetal liver development, polynucleotides of the present invention may also be used for the diagnosis of developmental disorders. Further, polypeptides of the present invention may be used therapeutically to treat developmental liver disorders as well as to regulate hepatocyte and supporting cell growth following hepatectomy or to treat liver disorders.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hemangiopericytoma and liver disorders.

## 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 61

This gene is expressed primarily in schizophrenic frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: nervous system and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the frontal cortex and CNS expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for study, treatment and diagnosis of frontal cortex, neuro-degenerative and CNS disorders

## 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in human adrenal gland tumor, and to a lesser extent in human kidney, medulla and adult pulmonary tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic, endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and nervous system disorders and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, kidney, brain and other tissue of the nervous system, pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of neurological and endocrine disorders including neoplasia.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 63

This gene is expressed primarily in human adipocytes, and to a lesser extent in spleen, 12-week old human, and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune, metabolic and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes, spleen, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of immune, developmental and metabolic disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

One translated product of this clone is homologous to the mouse zinc finger protein PZF. (See Accession No. 453376; see also Gene 152 (2), 233-238 (1995).) Preferred polypeptide fragments correspond to the highly conserved domains shared between mouse and man. For example, preferred polypeptide fragments comprise the amino acid sequence: LQCEICGFTCRQKASLNWHMKKHADASFYQFSCNICGKFEKDSVVAHKAKSH PEV (SEQ ID NO: 621); ITSTDILQTNPESLTOPSD (SEQ ID NO: 622); NSTSGECLLEAEGM SKSY (SEQ ID NO: 623); CSGTERVSLMADGKIFVGSQSGQTGELVMNSDILGATTEVLIEDSD SAGP (SEQ ID NO: 624); IQYVRCMEGCGTVLAHPRYLQHIIKYQHLLKKYVCPHPSCGRLF RLQKQLRHAKHHT (SEQ ID NO: 625); DQRDYICEYCARAFKSSHNLAVHRMIHTGEK (SEQ ID NO: 626); RSSRTSVSRHRDTEHTSSRSKTGSLQICKSEPTDQLDY (SEQ ID NO: 627); PFKDDPDETYKPHLERETPKRRKSG (SEQ ID NO: 630); QYVRCMEGCGTVLAHPRYLQ HHKYQHLLKKYVCPHPSCGRLFRLQKQLRHAKHHTD (SEQ ID NO: 629); or residues 151-182 of QRDYICEYCARAFKSSHNLAVHRMIHTGEKY (SEQ ID NO: 628). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Rhabdomyosarcoma, melanocyte and colon cancer tissue and to a lesser extent in smooth muscle, pancreatic tumor, and apoptotic T-cells.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hemopoietic, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striated muscle, melanocytes, colon, smooth muscle, pancreas, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of cancer and hemopoietic disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 65**

This gene is expressed primarily in human adipose and salivary gland tissue and to a lesser extent in human bone marrow and fecal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose, lower levels may be routinely detected in certain tissues and wounded tissues) or salivary gland, bone marrow, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis of metabolic and immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 66**

This translated product of this gene was recently identified as oxytocinase splice variant 1. (See Accession Nos. 2209276 and d1010078.) Preferred polypeptide fragments comprise the amino acid sequence: EMFDSLSTYFKGSSLLMLKTYLSEDFVQHAVVLYLHN HSYASIQSDDLWDSFNEVTNQLDYKRMKMTWTLQKGFPLVTVQKKGKELFIQGERFPLNMK PEIQSDTRRYM (SEQ ID NO: 631). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 67**

20 This gene is expressed primarily in hemopoietic cells, particularly apoptotic T-cells, and to lesser extent in primary dendritic cells and adipose tissue.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of apoptotic T-cells, primary dendritic cells, and adipose tissue present in a biological sample and for diagnosis of diseases and conditions: hemopoietic diseases including cancer and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the oral and intestinal mucosa as well as hemopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment of diseases of the immune system, including cancer, hemopoietic and infectious diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 68

This gene is expressed primarily in kidney cortex and to a lesser extent in infant brain, heart, uterus, and blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of kidney tissue present in a biological sample and for diagnosis of diseases and conditions: soft tissue cancer, inflammation, kidney fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrines systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, brain, and other nervous tissue, heart, uterus, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and fibroses.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 69

The translation product of this gene shares strong sequence homology with vertebrate and invertebrate protein tyrosine phosphatases.

This gene is expressed primarily in endometrial tumors, melanocytes, myeloid progenitors and to a lesser extent in infant brain, adipocytes, and several hematopoietic stem cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of transformed hematopoietic and epithelial cells present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of skin and endometrium, leukemia. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, melanocytes, bone marrow, adipocytes, hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and sequence similarity with tyrosine phosphatases indicate that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and hematopoietic disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 70

This gene is expressed primarily in osteoclastoma, breast, and infant brain and to a lesser extent in various fetal and transformed bone, ovarian, and neuronal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: degenerative conditions of the brain and skeleton. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of degenerative, neurological and skeletal disorders.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene was originally cloned from tumor cell lines. Recently another group has also cloned this gene, calling it the human malignant melanoma metastasis-suppressor (KISS-1) gene. (See Accession No. U43527.) Preferred polypeptide fragments comprise the amino acid sequence: LEKVASVGNRPTGQQLSGLLA (SEQ ID NO: 632); VHRREASCYQCAEPSGDL (SEQ ID NO: 633); RPALRQAGGTRPRQRVAGL (SEQ ID NO: 634); and AVNFRPQRSQM (SEQ ID NO: 635). Any frame shifts can easily be resolved using known molecular biology techniques.

This gene is expressed primarily in many types of carcinomas and to a lesser extent in many normal organs.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanomas, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of transformed organ tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. As a tumor suppressor gene, increase amounts of the polypeptide can be used to treat patients having a particular cancer.

25 The tissue distribution indicates that this gene and the translated product is useful for diagnosing and study of cancer.

## 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene is expressed primarily in striatum and to a lesser extent in adipocytes and hemangiopericytoma.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of striatal cells present in a biological sample and for diagnosis of diseases and conditions: neurological, fat and lysosomal storage

diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striatal tissue, adipocytes, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of neurodegenerative and growth disorders.

## 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

This gene is expressed primarily in bone marrow stromal cells and to a lesser extent in smooth muscle, testes, endothelium, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of bone marrow present in a biological sample and for diagnosis of diseases and conditions: connective tissue and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, stromal cells, smooth muscle, testes and other reproductive tissue, endothelium, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of connective tissue and blood diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 74**

This gene is expressed primarily in brain, fetal liver and lung and to a lesser extent in retina, spinal chord, activated T-cells and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of brain and regenerating liver present in a biological sample and for diagnosis of diseases and conditions: CNS and spinal chord injuries, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, pulmonary tissue, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of hematopoietic and neurological conditions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 75**

The translation product of this gene shares sequence homology with GTP binding proteins (intracellular).

This gene is expressed primarily in bone marrow, brain, and melanocytes and to a lesser extent in various endocrine and hematopoietic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic and nervous system conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, melanocytes, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to nucleotide binding factors indicate that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of brain degenerative, skin and blood diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 76**

This gene is expressed primarily in activated T-cells and to a lesser extent in retina, brain, and fetal bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of activated T-cells and developing brain present in a biological sample and for diagnosis of diseases and conditions: immune deficiencies and skeletal and neuronal growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and skeletomuscular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, retinal tissue, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of cancer, urogenital, and brain degenerative diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 77**

This gene is expressed primarily in fetal liver, activated monocytes, osteoblasts and to a lesser extent in synovial, brain, and lymphoid tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of myeloid and lymphoid present in a biological sample and for diagnosis of diseases and conditions: inflammation, immune deficiencies, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skeleton, expression of this gene at significantly



higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, blood cells, bone, synovial tissue, brain and other tissue of the nervous system, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of lymphoid and mesenchymal cancers and nervous system diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 78

The translation product of this gene shares sequence homology with polymerase polypeptide precursor which is thought to be important in DNA repair and replication

This gene is expressed primarily in infant brain and to a lesser extent in tumors and tumor cell lines

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, especially of the neural system and developing organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to polymerase polypeptide precursor indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers especially of the neural system and developing organs

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed primarily in muscle and endothelial cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., muscle, endothelial cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the vascular and neural system including cardiovascular and endothelial.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 80

This gene is expressed primarily in placenta and to a lesser extent in fetal liver

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental disorders and disorder of the haemopoietic system, fetal liver and placenta. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developmental disorders and disorder of the haemopoietic system, fetal liver and placenta, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of developmental disorders and disorders of the haemopoietic system, fetal liver and placenta.

# FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed primarily in bone marrow, placenta and tissues and organs of the hematopoietic system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the bone and haemopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, bone and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, placenta, and hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the immune, bone and hematopoietic system

# FEATURES OF PROTEIN ENCODED BY GENE NO: 82

The translation product of this gene shares sequence homology with secretory carrier membrane protein which is thought to be important in protein transport and export. Any frame shifts in coding sequence can be easily resolved using standard molecular biology techniques. Another group recently cloned this gene, calling it SCAMP. (See Accession No. 2232243.)

This gene is expressed primarily in prostate, breast and spleen, and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the breast prostate and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly disorders of the breast prostate and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., prostate, mammary tissue, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secretory carrier membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the breast, prostate and spleen.

# FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed primarily in developing organs and tissue like placenta and infant brain and to a lesser extent in developed organs and tissue like cerebellum and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, heart, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the neural system including neurological disorders and cancer.

# FEATURES OF PROTEIN ENCODED BY GENE NO: 84

The translation product of this gene shares sequence homology with ATPase 6 in Trypanosoma brucei which is thought to be important in metabolism.

This gene is expressed primarily in tumor and fetal tissues and to a lesser extent in melanocytes, kidney cortex, monocytes and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: metabolism disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, melanocytes, kidney, blood cells, ovary and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATPase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of metabolism disorders, especially in fetal and tumor tissue growth.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene shares sequence homology with the immunoglobulin superfamily of proteins which are known to be important in immune response and immunity.

20 This gene is expressed primarily in stromal cells, colon cancer, lung, amygdala, melanocyte and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells, colon, lung, amygdala, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution and homology to immunoglobulin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune system disorders.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 86

The translation product of this gene shares sequence homology with transcription initiation factor eIF-4 gamma which is thought to be important in gene transcription.

This gene is expressed primarily in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis.

10 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution and homology to transcription initiation factor eIF-4 gamma indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene regulation in tumorigenesis.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 87

The translation product of this gene shares sequence homology at low level in prolines with secreted basic proline-rich peptide IL-2 which is thought to be important in protein structure or inhibiting hydroxyapatite formation in vitro.

This gene is expressed primarily in endometrial tumor and fetal lung.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: endometrial tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular/skeletal and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 The tissue distribution and homology to secreted basic proline-rich peptide II-2 indicate that polynucleotides and polypeptides corresponding to this gene are useful for inhibiting hydroxyapatite formation or establishing cell/tissue structure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 88

- 10 This gene is expressed primarily in: amniotic cells induced with TNF in culture; and to a lesser extent in colon tissue from a patient with Crohn's Disease; parathyroid tumor; activated T-cells; cells of the human Caco-2 cell line; adenocarcinoma; colon; corpus colosum; fetal kidney; pancreas tumor; fetal brain; early stage brain, and anergic T-cells.

- 15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system; e.g., tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., amniotic cells, colon, kidney, pancreas, parathyroid, brain and other tissue of the nervous system, blood cells, hematopoietic cells, liver, spleen, bone, testes and other reproductive tissue, brain and other tissue of the nervous system, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30 The tissue distribution indicates that the protein product of this clone is useful for modulating tumorigenesis and other immune system conditions such as disorders in immune response.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 89

- 35 This gene is expressed primarily in fetal liver/spleen and hematopoietic cells and to a lesser extent in brain, osteosarcoma, and testis tumor.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

- biological sample and for diagnosis of diseases and conditions: leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, liver, spleen, bone, testes, and other reproductive tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 90

- 15 The translation product of this gene shares weak sequence homology with mouse Gcap 1 protein which is developmentally regulated in brain.

- 20 This gene is expressed primarily in infant and adult brain and fetal liver/spleen and to a lesser extent in smooth muscle, T cells, and a variety of other tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, spleen, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and its homology to Ccapi1 protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosis of disorders in neuronal, hematopoietic, immune, and endocrine systems.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in brain and hematopoietic cells and to a lesser extent in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorder in nervous, hematopoietic, immune systems and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the in nervous, hematopoietic, immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of disorders in the nervous, hematopoietic, and immune systems.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The translation product of this gene shares sequence homology with neuroendocrine-specific protein A which is thought to be important in neurologic systems.

This gene is expressed primarily in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neural disorders and degeneration disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central or peripheral nervous systems, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to neuroendocrine-specific protein A indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neural disorders and degeneration disease.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

The translation product of this gene shares sequence homology with collagen-like protein and prolin-rich protein which are thought to be important in connective tissue function and tissue structure.

This gene is expressed primarily in fetal liver/spleen and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to collagen-like protein and proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for supporting brain and hematopoietic tissue function and diagnosis and treatment of disorders in these functions.

## 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

This gene is expressed primarily in embryonic tissues and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to,. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system (e.g., tumors), expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in brain tumor, placenta, and melanoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor or melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain or melanocytes, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, placenta, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the translation product of this gene is useful in the diagnosis and treatment of brain tumors and melanoma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 96

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

This gene is expressed primarily in fetal liver; and to a lesser extent in placenta and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of fetal liver or defects of glucose-regulated ATPase activities in tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune/hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, placenta, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to yeast SUR4 membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of defects of fetal liver or defects of glucose-regulated ATPase activities.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 97

This gene is expressed primarily in fetal liver, brain, and amniotic fluid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the fetal immune system and adult brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune system and adult brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the fetal immune and hematopoietic systems since fetal liver is

the predominant organ responsible for hematopoiesis in the fetus. In addition, the gene product of this gene is thought to be useful for detecting certain neurological defects of the brain.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

The translation product of this gene shares sequence homology with an yolk protein precursor, Vitellogenin which is thought to be important in binding lipids such as phospholipids.

This gene is expressed primarily in amniotic cells and fetal liver.

- 10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in amniotic cells, fetal liver development and the fetal immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 25 The tissue distribution and homology to vitellogenin indicate that the protein product of this clone is useful for treatment and diagnosis of defects in amniotic cells, fetal liver development and the fetal immune system.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 99

- 30 This gene is expressed primarily in placenta, endometrial tumor, osteosarcoma and stromal cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumor of the endometrium or bone, and osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the obstetric system (e.g., placenta,

- endometrium) and the bones, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endometrium, bone, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors and abnormalities of the endometrium, and the bones because of its abundance in the aforementioned tissues.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene is expressed primarily in hepatocellular tumor.

- 15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatocellular tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 25 The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of hepatocellular cancer because of its abundant expression in this tissue.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 101

- 30 This gene is expressed primarily in Corpus Colosum, fetal lung and infant brain.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the Corpus Colosum or defects of the fetal lung. Similarly, polypeptides and antibodies directed to



these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Corpus Colosum and brain in general, and fetal lung, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of the Corpus Colosum and brain in general, and defects of fetal lung.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 102

This gene is expressed primarily in T cells and stromal cells, and to a lesser extent in adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of T cell immunity and stromal cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, stromal cells, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of T cell immunity and stromal cell development because of its abundant expression in these tissues.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 103

This gene is expressed primarily in infant brain and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, especially brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the brain, especially in young children.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 105

This gene is expressed primarily in human osteoclastoma and to a lesser extent in human pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly osteoclastoma and pancreatic tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in transformed tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone and pancreas, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of some types of tumors, particularly pancreatic cancer and osteoclastoma.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 106**

This gene is expressed primarily in fetal liver/spleen, and to a lesser extent in activated T-Cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of immune disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 107**

This gene is expressed primarily in human embryo and to a lesser extent in spleen and chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the diagnosis and treatment of leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 108**

This gene is expressed primarily in placenta, and to a lesser extent in early stage human brain and in lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: fetal developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in fetal and amniotic tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this is useful for production of growth factor(s) associated with fetal development. Preferred polypeptides comprise the full-length polypeptide shown in the sequence listing, truncated however, at the amino terminus and beginning with QTLE.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 109**

This gene is expressed primarily in fetal spleen, and to a lesser extent in B-Cell lymphoma and T-Cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the treatment and diagnosis of human lymphomas.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 110

5 The translation product of this gene shares sequence homology with sarcoma amplified sequence (SAS), a tetraspan receptor which is thought to be important in malignant fibrous histiocytoma and liposarcoma.

This gene is expressed primarily in human osteoclastoma, and to a lesser extent in pineal gland and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: malignant fibrous histiocytoma and liposarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, pineal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to sarcoma amplified sequence (SAS) indicate that the protein product of this clone is useful for treatment of, osteosarcoma, malignant fibrous histiocytoma and liposarcoma and related cancers, particularly sarcomas.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 111

The translation product of this gene shares sequence homology with 6.8K

30 proteolipid protein, mitochondrial - bovine.

This gene is expressed primarily in Wilm's tumor and to a lesser extent in cerebellum and placenta.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Wilm's tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the immune or renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to 6.8K proteolipid protein indicate that the protein product of this clone is useful for diagnostic and therapeutics associated with tumors, particularly Wilm's tumor disease.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 112

15 This gene is expressed primarily in embryonic tissue and to a lesser extent in osteoblasts, endothelial cells, macrophages (GM-CSF treated), and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, bone, endothelial cells, blood cells and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune disorders. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: MITDVQLAIFANMLGVSLFLVLYHYVA VNNPKKQE (SEQ ID NO: 636).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 113

35 This gene is expressed primarily in hepatocellular tumor, and to a lesser extent in fetal liver/spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors, particularly hepatocellular tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of tumors, particularly hepatocellular tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene exhibits a very high degree of sequence identity with the human Ptg8 gene which is thought to be important in p53 mediated apoptosis. The sequence of this gene has since been published by Polyak and colleagues (Nature 389, 300-306 (1997)). In addition, the predicted translation product of this contig exhibits very high sequence homology with a murine gene denoted as El24 which is also thought to be important in p53 mediated apoptosis.

This gene is expressed primarily in infant brain and activated T-cells and to a lesser extent in bone marrow, fetal liver, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and tissue damage by radiation and anti-cancer drugs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, bone marrow, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human Ptg8 and murine El24 genes indicate that polynucleotides and polypeptides corresponding to this gene are useful for preventing apoptosis in patients being treated with anti-oncogenic drugs such as etoposide, hydroperoxycyclophosphamide, and X-irradiation, since this protein product is upregulated in cells undergoing such treatment where p53 was overexpressed. It may also be useful in the treatment of hematopoietic disorders and in boosting numbers of hematopoietic stem cells by interfering with the apoptosis of progenitor cells. The mature polypeptide is predicted to comprise the following amino acid sequence:

EEMADSVKTFQLDARGIKDSIWQICTSKLDARIQKREQRRASSVLAQRRASQIERKQES  
EPRIVSKRQCCAVNCGVWFPSLLFVRFVPLQSVTARIGDPSLHGDDVWSWLEFLTISFSA  
LWVLPFLVLSKVYNAVWFQIADLAFEVSGKKPHFPFSVSKHADMFLNLLQALFLQGMFVSL  
FPIHLVQQLVSLHMSLLVSLYCFEYRWRNKGIEMHQRLSNIERNWPTYRFQFGLPLAFLTAMQ  
SSYISGCLFSLFPLISANEAKTPGKAYLQRLFSLVFLSNRLEFKTVYVLOSALSSSTSAREK  
FSPHPSPARKLKATAGH (SEQ ID NO: 637). Accordingly, polypeptides comprising the foregoing amino acid sequence are provided as are polynucleotides encoded such polypeptides.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 115

This gene is expressed primarily in stromal cells and to a lesser extent in multiple sclerosis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: affecting the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of multiple sclerosis and other autoimmune diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 116**

This gene is expressed primarily in the gall bladder

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: gall stones or infection of the digestive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for possible prevention of digestive disorders where there may be a lack of digestive enzymes produced or in the detection and possible prevention of gall stones.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 117**

The translation product of this gene shares sequence homology with dystrophin gene which is thought to be important in building and maintenance of muscles.

This gene is expressed primarily in placenta and to a lesser extent in fetal brain and fetal liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: muscular dystrophy, Duchenne and Becker's muscular dystrophies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal muscle system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, muscle, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the dystrophin gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diseases related the degenerative myopathies that are characterized by the weakness and atrophy of muscles without neural degradation, such as Duchenne and Becker's muscular dystrophies.

**10 FEATURES OF PROTEIN ENCODED BY GENE NO: 118**

This gene is expressed primarily in olfactory tissue and to a lesser extent in cartilage.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: connective tissue diseases; chondrosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the connective tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue and cartilage, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for tumors of connective tissues, osteoarthritis and the treatment and diagnosis of chondrosarcoma.

**30 FEATURES OF PROTEIN ENCODED BY GENE NO: 119**

This gene is expressed primarily in Activated Neutrophils and to a lesser extent in fetal spleen, and CD34 positive cells from cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergies, defects in hematopoiesis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoiesis system the, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

10 corresponding to this gene are useful for reducing the allergic effects felt by allergy suffers by neutralizing the activity of the immune system, especially since neutrophils are abundant in persons suffering from allergies and other inflammatory conditions.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 120

The translation product of this gene shares sequence homology with poly A binding protein II which is thought to be important in RNA binding for transcription of RNA to DNA

This gene is expressed primarily in colon and to a lesser extent in brain and immune system.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: colon cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, tissue and cells of the immune system, and brain or other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution and homology to poly A binding protein II indicate that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of colon cancer and other disorders of the digestive system..

#### 1 FEATURES OF PROTEIN ENCODED BY GENE NO: 121

The translation product of this gene shares sequence homology with thymidine diphosphoglucose 4,6 dehydrase which is thought to be important in the metabolism of sugar.

5 This gene is expressed primarily in fetal liver and spleen and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diabetes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution and homology to thymidine diphosphoglucose 4,6 dehydrase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of persons with diabetes since it appears that this protein is needed in the metabolism of sugar in to its more basic components.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with ceruloplasmin which is thought to be important in the metabolism and transport of iron and copper. Ceruloplasmin also contains domains with homology to clotting factors V and VIII. Defects in the circulating levels of ceruloplasmin (aceruloplasminemia) have been associated with certain disease conditions such as Wilson disease, and the accompanying hepatolenticular degeneration.

30 This gene is expressed primarily in brain and retina and to a lesser extent in endothelial cells.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases marked by defects in iron metabolism; aceruloplasminemia not characterized by defects in the

known ceruloplasmin gene locus; nonclassical Wilson disease; movement disorders; and tumors derived from a brain tissue origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, retina, and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, retinal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ceruloplasmin indicate that

polynucleotides and polypeptides corresponding to this gene are useful for treatment of patients with aceruloplasminemia, or other defects in iron and/or copper metabolism. Mutations in this locus could also be diagnostic for patients currently experiencing or predicted to experience aceruloplasminemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene is expressed primarily in brain and B cell lymphoma and to a lesser extent in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: B cell lymphoma; tumors and diseases of the brain and/or spleen; hematopoietic defects. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in neuronal,

hematopoietic, and immune systems. It could potentially be useful for neurodegenerative disorders and neuronal and/or hematopoietic cell survival or proliferation.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in osteoclastoma, dermatofibrosarcoma, and B cell lymphoma and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer in particular osteoclastoma, dermatofibrosarcoma, and B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, immune, and circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, epidermis, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers and lymphoma; osteoporosis; and the control of cell proliferation and/or differentiation.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in immune tissues and hematopoietic cells, particularly in activated T cells and neutrophils, spleen, and fetal liver, and to a lesser extent in infant adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in T cell activation; hematopoietic disorders; tumors of a hematopoietic and/or adrenal gland origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and/or endocrine systems, expression of this gene at significantly higher



or lower levels may be routinely detected in certain tissues and cell types (e.g., cells and tissues of the immune system, hematopoietic cells, blood cells, liver, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and/or hematopoietic disorders; diseases related to proliferation and/or differentiation of hematopoietic cells; defects in T cell and neutrophil activation and responsiveness; and endocrine and/or metabolic disorders, particularly of early childhood.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in placenta and endothelial cells and to a lesser extent in melanocytes and embryonic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell origin; angiogenesis associated with tumor development and metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, melanocytes, and embryonic tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of developmental disorders; inhibition of angiogenesis; and vascular patterning.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in endothelial cells and hematopoietic tissues, including spleen, tonsils, leukocytes, and both B- and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell and/or hematopoietic origin; leukemias and lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, hematopoietic cells, spleen, tonsils, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the manipulation of angiogenesis; the differentiation and morphogenesis of endothelial cells; the proliferation and/or differentiation of hematopoietic cells; and the commitment of hematopoietic cells to distinct cell lineages.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 128

This gene is expressed primarily in kidney medulla and to a lesser extent in spleen from chronic myelogenous leukemia patients, prostate cancer, and some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a kidney origin; chronic myelogenous leukemia; prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, spleen, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of kidney disorders and cancer, particularly chronic myelogenous leukemia and prostate cancer. It may also be useful for the enhancement of kidney tubule regeneration in the treatment of acute renal failure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 129

This gene is expressed primarily in adult and infant brain and to a lesser extent in mesenchymal or fibroblast cells, as well as tissues with a mesenchymal origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; fibrosis. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and of mesenchymal cells and tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; and fibrosis, based upon the expression of this gene within those tissues. Fibrosis is considered as mesenchymal cells and fibroblasts are the primary cellular targets involved in this pathological condition.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 130

This gene is expressed primarily in hepatocellular cancer and to a lesser extent in fetal tissues as well as testes tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, fetal tissue, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 131

This gene is expressed only in infant early brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: development and diseases of the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the brain in children and in treating nervous system disorders such as Alzheimer's disease, schizophrenia, dementia, depression, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in brain and to a lesser extent in glioblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Alzheimer's disease,

schizophrenia, depression, mania, and dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating brain disorders such as Alzheimer's disease, schizophrenia, depression, mania, and dementia.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 133

The translation product of this gene shares sequence homology with ribitol dehydrogenase of bacteria which is thought to be important in metabolism of sugars.

This gene is expressed primarily in macrophage and to a lesser extent in T-cell lymphoma and lung.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tissue destruction in inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and homology to ribitol dehydrogenase indicate that polynucleotides and polypeptides corresponding to this gene are useful for altering macrophage metabolism in diseases such as inflammation where macrophages are causing excess tissue destruction.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 134

This gene is expressed primarily in pancreatic tumor and to a lesser extent in synovial sarcoma.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and connective tissue systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pancreas, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing various cancers.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 135

This gene is expressed primarily in T cell lines such as Raji and to a lesser extent in infant brain.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune system disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing inflammatory diseases

such as rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, as well as neutropenia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 136

The translation product of this gene shares high sequence homology with SAR1 subfamily of GTP-binding proteins which is thought to be important in vesicular transport in mammalian cells.

This gene is expressed primarily in serum-stimulated smooth muscle cells and to a lesser extent in a T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases affecting vesicular transport. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP-binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene therapy in treating the large number of diseases involved in defective vesicular transport within cells..

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 137

The translation product of this gene shares sequence homology with a protein found in *C. elegans* cosmid F25B5.

This gene is expressed primarily in a fetal tissues and to a lesser extent in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development, especially of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal pulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, pulmonary tissue, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment and diagnosis of diseases affecting the pulmonary system, such as emphysema.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene is expressed primarily in gall bladder and to a lesser extent in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: digestive system disease and gall bladder problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the digestive system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene is expressed primarily in placenta and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developing tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing abnormal fetal development.

#### FEATURES OF PROTEIN ENCODED BY GENE NO. 140

This gene is expressed primarily in smooth muscle and to a lesser extent in ovary, prostate cancer, and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hypertension and atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,

particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, ovary and other reproductive tissue, prostate, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the circulatory system, such as hypertension, atherosclerosis, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO. 141

This gene is expressed primarily in fetal spleen and to a lesser extent in placenta and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and other diseases affecting blood cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen, placenta, bone marrow, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the generation of red and white blood cells and for the diagnosis of disease of these cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO. 142

The predicted translation product of this contig is a human homolog of the murine tetracycline/sugar transporter molecule recently reported by Matsuo and colleagues (Biochem. Biophys. Res. Commun. 238 (1), 126-129 (1997)).

This gene is expressed primarily in synovium and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: rheumatoid arthritis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and lymphatic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, such as rheumatoid arthritis, leukemia, neutropenia, inflammatory bowel disease, psoriasis, sepsis, and the like.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 143

This gene is expressed primarily in placenta and to a lesser extent in melanocyte, fetal liver and spleen, and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal early development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, melanocytes, liver, spleen, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of abnormal early development phenomena and diseases.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and neutropenia.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and blood systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the

35

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in hematopoiesis and bone marrow regeneration as it is most abundant in fetal tissues responsible for the generation of hematopoietic cells.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 145

The translation product of this gene shares sequence homology with protein tyrosine phosphatase which is thought to be important in transducing signal to activate cells such as T cell, B cell and other cell types.

This gene is expressed primarily in T cells and tissues in early stages of development and to a lesser extent in cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic and fetal tissue, undifferentiated cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution and homology to the protein tyrosine phosphatase family indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating the immune system.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 146

This gene is expressed primarily in T cell and to a lesser extent in B cell, macrophages and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

35

providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

10 corresponding to this gene are useful for regulating the immune system therefore can be used in treating diseases such as autoimmune diseases and cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 147

This gene is expressed primarily in placenta and to a lesser extent in endothelial cells, testis tumor, ovarian cancer, uterine cancer.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, testis and ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This sequence has significant homology to mouse torsin A. Recently, another group cloned the human Torsin A gene. (See, Accession No. Z358279; see also Nature Genet. 17, 40-48 (1997).)

35 This gene is expressed primarily in osteoclastoma, T-cell, and placenta and to a lesser extent in fetal lung, fetal liver, fetal brain, adult brain and tumor tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions; disease conditions in hematopoiesis and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, bone, placenta, lung, liver, and brain and other tissues of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating blood related diseases such as deficiencies in red blood cell, white blood cell, platelet and other hematopoiesis cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 149

20 This gene is expressed primarily in T cell, prostate and prostate cancer, endothelial cells and to a lesser extent in monocyte, dendritic cell, bone marrow, salivary gland, colon cancer, stomach cancer, pancreatic tumor, uterine cancer, fetal spleen and osteoclastoma.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions; immuno-related diseases and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, prostate, endothelial cells, dendritic cells, bone marrow, salivary gland, colon, stomach, pancreas, uterus, spleen and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.



The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 150

5 This gene was recently cloned by another group, calling it cIF3-p66. (See Accession No. 2351378.) This gene plays a role in RNA binding and macromolecular assembly, and therefore, any mutations in this gene would likely result in a diseased phenotype. Preferred polypeptide fragments comprise the amino acid sequence:

10 MAKFMTPVIQDNPSQWGPCAVPEQFRDMPYQPFSGKDRLGKVDWTGATYQDKRYTNKYSS  
QFGGGSQYAYFHEDESSFLVDARTQKTAYQRNRMFAQRNLRDRDRNNMLQFNLQLP  
KSAKQKERIRLQKKFQKQGVQRKWDQKSKPRDSSVEVRSDWEVKEEMDFPQLMKMRY  
LEVSEPDIECCGALEYDKAFDRITRSEKPLRXXKRUFHTVTTDDPVRKLAKTQGNVFAITD  
AILATLMSCTRSVYSWDIVVQVSGSLFFDKRDNSDFDLLTVSETANEPPQDEGNSFNSPRL  
AMEATYINHNSQQCLRMGKERYNFPNPNPFVEDMDKNEIASVAYRYSCKLGDDIDLIVRC  
15 EHDGVMGTANGEVFINIKTLNEWSRHCNGVDWRQLDSQGA VIATELKNNSYKLARWTC  
CALLAGSEYLKLGYYRYHVHVDSSRHVILGTQQFKPNEFASQINTSVENAWGILRCVIDICMKL  
EEGKYLILKDPNKQVIRVYSLPDCTFSS (SEQ ID NO: 638), as well as N-terminal and C-terminal deletions of this polypeptide fragment.

This gene is expressed primarily in T cell, bone marrow, embryo and endothelial cells and to a lesser extent in testis tumor and endometrial tumor.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders and cancers.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 151

This gene is expressed primarily in testis and to a lesser extent in T cell, spinal cord, placenta, neutrophil and monocyte.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive tissue, blood cells, tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell  
10 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating immune and reproductive functions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 152

The translation product of this gene shares sequence homology with tyrosyl-tRNA synthetase which is thought to be important in cell growth.

25 This gene is expressed primarily in brain, liver, keratinocytes, tonsils, and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, keratinocytes, tonsils, heart  
30 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissues of the nervous system, liver, keratinocytes, tonsils and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to tyrosyl-tRNA synthetase indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 153

This gene is homologous to the *Drosophila* transcriptional regulator *dre4*. (See Accession No. 2511745.) *Dre4* is a gene required for steroidogenesis in *Drosophila* melanogaster and encodes a developmentally expressed homologue of the yeast transcriptional regulator CDC68. Preferred polypeptide fragments comprise the amino acid sequence: KKRHTDVQFTTEVGETTDLGKHQMHDRDLYAEQMEREMRHKLTAFKN FLEKVEALTKBELFEFVFRDLGNQAPYRSTCLQPTSSALVNATEWPPVVTLDEVELHFEXR VQFHLKNPDMVIVYKDYSKKVTMAINAPVASLDPIKEWLNQCDLKYTEGVQSLNWTMKMTVD DPEGFEEQGQWSFL (SEQ ID NO: 639), as well as N-terminal and C-terminal deletions of this fragments. Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in fetal liver, spleen, placenta, lung, T cell, thyroid, testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor, heart and liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal liver, spleen, placenta, lung, T cell, thyroid, testes expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, lung, blood cells, thyroid, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed primarily in brain and to a lesser extent in fetal heart, testis, spleen, lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: heart, liver and spleen diseases, immunological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, fetal heart, testis, spleen, lung expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, heart, testes and other reproductive tissue, spleen, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 155

Activation of T cells through the T cell antigen receptor (TCR) results in the rapid tyrosine phosphorylation of a number of cellular proteins, one of the earliest being a 100 kDa protein. This gene is the human equivalent of murine valosin containing protein (VCP). VCP is a member of a family of ATP binding, homo-oligomeric proteins, and the mammalian homolog of *Saccharomyces cerevisiae* cdc48p, a protein essential to the completion of mitosis in yeast. Both endogenous and expressed murine VCP are tyrosine phosphorylated in response to T cell activation. Thus we have identified a novel component of the TCR mediated tyrosine kinase activation pathway that may provide a link between TCR activation and cell cycle control.

This gene is expressed primarily in brain, liver, spleen, placenta. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, spleen, placenta expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, spleen, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to VCR indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with rat growth response protein which is thought to be important in cell growth. A group recently cloned the human homolog of this gene, calling it insulin induced protein 1. (See Accession No. 2358269, see also, Genomics 43 (3), 278-284 (1997).) Preferred polypeptide fragments comprise the amino acid sequence: RSGLGICITATLITQF LVYNGVYQYTSPDFLYIRSWLPCIFFSGGVTVGNIGRQLAMGVPEKPHSD (SEQ ID NO: 640), as well as N-terminal and C-terminal deletions of this polypeptide fragment. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain, liver, placenta, heart, spleen, lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, placenta, heart, spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, placenta, heart, spleen, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to growth-response protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 157

This gene is expressed primarily in Glioblastoma, endometrial tumor, lymphoma and pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Glioblastoma, Endometrial tumor, lymphoma and pancreas tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, lymphoid tissue, pancreas, and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 158

The translation product of this gene shares sequence homology with ICE receptor which is thought to be important in allergy and asthma.

This gene is expressed primarily in T cell, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergy and asthma and other immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to IgE receptor indicate that polynucleotides and polypeptides corresponding to this gene are useful for allergy and asthma.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 159

The translation product of this gene shares sequence homology with immunoglobulin heavy chain which is thought to be important in immune response to the antigen.

This gene is expressed primarily in activated neutrophil and to a lesser extent in activated T cell, monocyte and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: infection, inflammation and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin heavy chain variable region indicate that polynucleotides and polypeptides corresponding to this gene are useful for making the ligand to block specific antigen which cause certain disease.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with mouse X inactive specific transcript protein which is thought to be important in X chromosome inactivation.

This gene is expressed primarily in HSA172 cell and to a lesser extent in normal ovary tissue, ovarian cancer, frontal cortex and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian tumor, schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to X inactive specific transcript protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive system tumors and CNS tumors.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene is expressed primarily in adipose cell and to a lesser extent in liver and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity and liver disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose cell, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose cells, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of obesity and liver disorder.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 162

The translation product of this gene shares sequence homology with yeast ubiquitin activating enzyme homolog which is thought to be important in protein posttranslational processing.

This gene is expressed primarily in stromal cell and to a lesser extent in retina, H. Atrophic Endometrium, colon carcinoma and myeloid progenitor cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development, neuronal growth disorders and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal cells, endometrium, colon, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ubiquitin-activating enzyme homolog indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of some type of tumors, fucosidosis and neuronal growth disorders.

20

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in primary breast cancer and hemangiopericytoma and to a lesser extent in adult brain and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer, leukemia and cerebellum disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of various tumors and disease involved in neural system.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 164

The translation product of this gene shares sequence homology with proline rich proteins. Recently, another group has also cloned this gene, calling it CD84 leukocyte antigen, a new member of the Ig superfamily. (See Accession No. U82988, see also, Blood 90 (6), 2398-2405 (1997).)

10 This gene is expressed primarily in Weizmann olfactory tissue and osteoclastoma and to a lesser extent in anergic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ostsis and immune disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue, bone, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution and homology to the Ig superfamily indicate that the protein product of this clone is useful for treatment of osteoporosis, autoimmune disease, and other immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 165

30 This gene is expressed primarily in atrophic endometrium and colon cancer and to a lesser extent in some fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system,

35

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, colon, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors, specifically endometrium and colon tumors.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human primary breast cancer and to a lesser extent in activated monocyte. Although the predicted signal sequence is identified in Table 1, other upstream sequences are also relevant. Preferred polypeptide fragments comprise the amino acid sequence: VTQPKHLASMGVPEIPSPFYVWELAXXPXVRISWRGHFHG QSPYSTRPSIHKDYVNRLELNVTEQESQFLRISNLKEDQSVYFCRVELDRSG (SEQ ID NO: 641), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of breast cancer.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 167

This gene is expressed primarily in fetal tissues and to a lesser extent in adult lung. This gene has also been mapped to chromosomal location 9q34, and thus, can be used as a marker for linkage analysis for chromosome 9.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryo tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 168

The translation product of this gene shares sequence homology with Ig Heavy Chain which is thought to be important in immune response.

This gene is expressed primarily in prostate cancer tissue specifically. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The translation product of this gene shares sequence homology with cytosolic acyl coenzyme-A hydrolase, which is thought to be important in neuron-specific fatty acid metabolism. The gene represented by this contig has since been published by Hajra and colleagues (GenBank Accession No. U91316).

This gene is expressed primarily in human pituitary gland and to a lesser extent in colorectal cancer tissue. This gene has also been observed in the LNCAP cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hyperlipidemias of familial and/or idiopathic origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly blood, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to rat cytosolic acyl coenzyme-A

hydrolase indicate that polynucleotides and polypeptides corresponding to this gene are useful for the detection or treatment of hyperlipidemia disease states by virtue of the ability of specific drugs to activate the enzyme.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 170

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene which is thought to be important in organism development.

This gene is expressed primarily in human synovial sarcoma tissue, bone marrow, and to a lesser extent in human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of bone, specifically synovial sarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, connective tissues and possibly immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, bone marrow, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to *Caenorhabditis elegans* indicate that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic and/or therapeutic modality directed at the detection and/or treatment of connective tissue sarcomas or other related bone diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 171

The translation product of this gene shares sequence homology with beta1-6GlcNAc transferase which is thought to be important in the transfer and metabolism of beta1-6, N-acetylglucosamine. This gene product has previously been shown to suppress melanoma lung metastasis in both syngeneic and nude mice, decreased invasiveness into the matrigel, and inhibition of cell attachment to collagen and laminin without affecting cell growth.

This gene is expressed primarily in human testes and prostate tissues, and to a lesser extent in kidney, medulla, and pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, prostate, kidney, pancreas, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to beta1-6GlcNAc transferase indicate that the protein product of this clone is useful for the development of diagnostic and/or therapeutic modalities directed at the detection and/or treatment of cancer, the metastasis of malignant tissue or cells. Defects in this potentially secreted enzyme may play a role in metastasis.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 172**

This gene is expressed primarily in fetal spleen and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders, Wilms tumor disease, hepatic disorders, and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and identification of fetal defects along with correcting diseases that affect hematopoiesis and the immune system.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 173**

The translation product of this gene shares sequence homology with ret II

oncogene which is thought to be important in Hirschsprung disease and many types of cancers.

This gene is expressed in multiple tissues including the lymphatic system, brain, and thyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Hirschsprung disease and multiple cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, thyroid, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ret II oncogene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of various cancers. It would also be useful for the diagnosis and treatment of Hirschsprung disease. Preferred polypeptides of the invention comprise the amino acid sequence: MEAQVNEAESAREQLXLDIAGQKASKQELTELRLKQEFHYIEDLY RTKNTLQSRUKDRDEIQKLRLNQLTNKTLSNSQSELENRLHQLTETLIQKQTMLESISTEKNLS VQLERLQGNMSASQSSNGSSINMGIDNEGCTLRVVPVLENDITNLAGMYCKYRKAAS SIDQSRILGFLRRYPLRVFVIYMLLHLVVMIVLLTTPBMHHDQPYGK (SEQ ID NO: 642).

**FEATURES OF PROTEIN ENCODED BY GENE NO: 174**

The translation product of this gene shares sequence homology with testis enhanced gene transcript which is thought to be important in regulation of human development.

This gene is expressed primarily in infant brain and to a lesser extent in a variety of other tissues and cell types, including the prostate, testes, monocytes, macrophages, dendritic cells, keratinocytes, and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological, developmental, immune and inflammation disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, prostate, testes and other reproductive tissue, blood cells, keratinocytes, and adipocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to testis enhanced gene transcript indicate that the protein product of this clone is useful for diagnosis and treatment of disorders involving the developing brain and the immune system.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 175**

This gene is expressed primarily in prostate and to a lesser extent in various other tissues, including placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, especially of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of prostate disorders and cancer. It may also be useful for the diagnosis and treatment of endocrine disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 176**

The translation product of this gene shares sequence homology with *Saccharomyces cerevisiae* YNT20 gene which is thought to be important in mitochondrial function.

This gene is expressed at a particularly high level in muscle tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases related to such tissues and cell types including: muscle wasting diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., muscle and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the YNT20 gene indicate that this protein is useful for treatment and detection of neuromuscular diseases caused by loss of mitochondrial function. For example this gene or its protein product could be used in replacement therapy for such diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 177**

This gene is expressed primarily in the brain and to a lesser extent in kidney, placenta, smooth muscle, heart and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuromuscular diseases, degenerative diseases of the central nervous system, and heart disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, central nervous system, and heart, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, kidney, placenta, muscle, heart and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

This gene or its protein product could also be used for replacement therapy for the above mentioned diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 178**

The translation product of this gene shares sequence homology with caldesmon which is thought to be important in the cellular response to changes in glucose levels.

This gene is expressed primarily in multiple tissues including brain and retina.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: central nervous system disorders and retinopathy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the CNS disorders and retinopathy, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and retinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to caldesmon indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of retinopathies.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 179

The translation product of this gene shares sequence homology with mouse fibrosin protein which is thought to be important in regulation of fibrinogenesis in certain chronic inflammatory diseases.

This gene is expressed primarily in amniotic cells and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of breast cancer and abnormal embryo

development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to fibrosin indicate that the protein product of this clone is useful for treatment of breast cancer. This gene or its protein product could be used in replacement therapy for breast cancer. In addition the protein product of this gene is useful in the treatment of chronic inflammatory diseases.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 180

This gene is expressed several infant tissues including brain and liver and various adult tissues including brain, lung, liver, testes, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, brain cancer, lung cancer, liver cancer and cancers of the reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, liver, testes and other reproductive tissue, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene product indicates that the protein product of this clone is involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in activated monocytes and to a lesser extent in melanocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of immune system diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, melanocytes, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 182**

This gene is expressed primarily in placenta and several tumors of various tissue origin and to a lesser extent in normal tissues including liver, lung, brain, and skin,

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers of all kinds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

10 of the above tissues or cells, particularly of the central nervous system, respiratory system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, lung, brain and other tissues of the nervous system, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

15 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The high expression of this gene in multiple tumors indicates that the protein product of the clone may be involved in cell growth control and therefore would be useful for treatment of certain cancers. Likewise molecules developed to block the

20 activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 183**

25 The translation product of this gene shares sequence homology with the mouse Ndr1 gene which is thought to be important in cancer progression.

This gene is expressed multiple cell types and tissues including brain, lung, kidney, bone marrow, liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

30 of the above tissues or cells, particularly of the nervous, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous

35 system, lung, kidney, bone marrow, liver and spleen, and cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to Ndr1 gene, which is thought to be involved in cancer progression, indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 184**

This gene is expressed primarily in early stage human brain and liver and to a lesser extent in several other fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain and liver cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

15 central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 185**

This gene is expressed primarily in infant and embryonic brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of degenerative nervous system disorders and brain cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful

35 in providing immunological probes for differential identification of the tissue(s) or cell

115

type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in multiple tissues including placenta, fetal lung, fetal liver, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers including liver, brain and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, pulmonary system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, lung, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

116

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	11	582	1	582	177	177	313	1	18	19	22
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	197	1020	296	830	442	442	499	1	18	19	22
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	12	465	1	465	81	81	314	1	30	31	128
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	198	524	229	343		196	500	1	20	21	33
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	13	474	1	474	1	1	315	1	24	25	28
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	199	332	1	319	35	35	501	1	24	25	28

[illegible][illegible]

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	23	1486	596	1418	102	102	325	1	54	55	252
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	203	847	1	839	87	87	505	1	30	31	75
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	204	852	75	850		690	506	1			10
13	HTXEF04	209235 09/04/97	Uni-ZAP XR	205	1354	54	1354	100	100	507	1	33	34	207
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	24	2323	1017	2059	1242	1242	326	1	21	22	68
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	206	1378	113	1226	303	303	508	1	25	26	36
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	25	683	1	683	304	304	327	1	30	31	84

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	207	1166	281	884	567	567	509	1	18	19	19
16	HHFFL33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	26	2036	14	1959	214	214	328	1	20	21	36
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	27	717	1	717	70	70	329	1	30	31	63
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	208	697	2	697	33	33	510	1	31	32	32
18	HMDAE90	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	28	495	1	495	39	39	330	1	24	25	35
19	HOUAW01	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	29	556	1	556	116	116	331	1	19	20	23





Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
28	HSSDM73	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	38	672	1	672	22	22	340	1	38	39	42
29	HBMVK68	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	39	1908	135	1908	309	309	341	1	20	21	26
30	HMKDC66	97898 02/26/97 209044 05/15/97	pSport1	40	458	93	458	147	147	342	1	24	25	26
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	41	1153	500	1153	427	427	343	1	30	31	157
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	213	1079	502	896		739	515	1	23	24	43
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	42	1983	1092	1983	27	27	344	1	11	12	520

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	214	3791	2757	3357		2030	516	1			3
33	HTOJN06	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	43	1406	1	695		19	345	1	19	20	39
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	44	1391	851	1153	74	74	346	1	30	31	234
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	215	1334	822	1036		638	517	1	18	19	174
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	45	1569	768	1569	14	14	347	1	19	20	169
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	216	1511	770	1404	844	844	518	1	32	33	43



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	220	1258	149	1190	254	254	522	1	18	19	26
45	HSKDK47	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	55	1896	596	1614	650	650	357	1	33	34	47
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	56	1753	555	1753	414	414	358	1	18	19	73
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	221	1693	554	1693		526	523	1	25	26	58
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	57	1220	690	1024	128	128	359	1	30	31	102
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	222	1196	712	1163		1097	524	1			19

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
48	HFCAl74	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	58	1049	362	1049	335	335	360	1	33	34	48
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	59	1776	854	1737	189	189	361	1	30	31	179
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	223	1791	979	1791	1164	1164	525	1	18	19	40
50	HLFBC91	97899 02/26/97 209045 05/15/97	pBluescript SK-	60	443	1	443	164	164	362	1	21	22	25
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	61	2888	1909	2888	90	90	363	1	30	31	224
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	224	2517	1597	2517	1953	1953	526	1	18	19	57



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	230	448	9	448		1	532	1	22	23	75
60	HHGCM20	97899 02/26/97 209045 05/15/97	Lambda ZAP II	70	245	1	245	93	93	372	1	1	2	51
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	71	361	1	361	1	1	373	1	30	31	61
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	231	407	1	407	210	210	533	1	17	18	60
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	72	713	8	713	169	169	374	1	30	31	40
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	232	830	190	580	329	329	534	1	28	29	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	73	862	1	862	67	67	375	1	30	31	44
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	233	932	138	905	287	287	535	1			2
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	74	4602	4162	4525	730	730	376	1	30	31	203
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	234	2786	2406	2739	2577	2577	536	1	22	23	36
65	HSGBA84	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	75	1255	1	1195	112	112	377	1	28	29	29
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	76	475	1	475	13	13	378	1	30	31	136

Gene No.	cDNA Clone ID	ATCC Deposit No.: Z and Date	Vector	NT SEQ ID NO:	Total NT Seq.	5' NT B. NT	5' NT of Clone Seq.	Codon Start	5' NT of First AA	AA SEQ ID NO:	Y of AA	First AA	Sig of AA	Pep of AA	Secreted of AA	Last ORF
71	HFFAH94	97900 02/26/97 209046 05/15/97	Lambda ZAP II	81	1290	768	1054	701	701	383	I	21	22	138		
72	HBIA195	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	82	684	I	684	119	119	384	I	30	31	74		
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	83	2024	1609	1953	200	200	385	I	30	31	521		
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	237	1286	391	959					9	10	11		
74	HEBEG68	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	84	931	14	537	85	85	386	I	25	26	137		
75	HBIBAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	85	825	59	802	66	66	387	I	30	31	186		

[illegible]



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	238	734	1	734	1	1	540	1	37	38	108
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	239	809	80	794		294	541	1	15	16	106
76	HTXDU73	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	86	1238	36	918	17	17	388	1			1
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	87	1460	9	1458	166	166	389	1	53	54	299
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	240	2201	841	2080	507	507	542	1	43	44	136
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	241	1661	311	1520	390	390	543	1	35	36	424

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
78	HTEIY30	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	88	1395	567	1395	639	639	390	1	36	37	49
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	89	1186	352	1186	540	540	391	1	49	50	61
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	242	1146	329	1146	564	564	544	1	21	22	39
80	HPMFL27	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	90	1821	1203	1614	1503	1503	392	1	30	31	79
81	HMWDN32	97900 02/26/97 209046 05/15/97	Uni-Zap XR	91	862	253	862	359	359	393	1	32	33	36
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	92	696	349	696	98	98	394	1	30	31	180

Gene No.	cDNA Clone ID	ATCC Deposit No.: Z and Date	Vector	NT SEQ ID NO:	Total NT Seq.	5' NT 3' NT Clone of Seq.	5' NT Start Codon	5' NT First AA of Signal Rep	AA SEQ ID NO:	Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted AA of Secreted Portion	Last AA of Secreted of ORF
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	243	1350	265	1230	348	348	545	I	32	33	58
83	HHEFW36	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	93	1886	I	1759	197	197	395	I			21
84	HE2PL77	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	94	1774	742	1772	785	785	396	I	21	22	60
85	HSDFV29	209076 05/22/97	Uni-ZAP XR	95	2503	I	1648	206	206	397	I	32	33	152
85	HCOAV53	97901 02/26/97 209047 05/15/97	Lambda ZAP II	244	1529	72	911	191	191	546	I	20	21	33
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	96	2801	418	2801	234	234	398	I	30	31	480
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	245	1537	I	1537	125	125	547	I	21	22	367

[illegible]

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
91	HTSEL31	97901 02/26/97 209047 05/15/97	pBluescript	101	1394	608	1346	602	602	403	1	23	24	87
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	102	794	1	794	518	518	404	1	30	31	92
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	248	1766	42	1766	356	356	550	1	30	31	168
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	249	2664	47	1708		147	551	1	18	19	124
93	HODAS59	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	103	1544	898	1531	975	975	405	1			21
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	104	871	106	871	248	248	406	1	34	35	174

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	250	865	97	865	258	258	552	1	19	20	177
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	105	404	1	404	16	16	407	1	21	22	64
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	251	2082	852	2074	829	829	553	1	22	23	72
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	106	1542	506	1542	122	122	408	1	51	52	280
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	252	1482	508	1482		633	554	1	15	16	45
96	HCUHB01	209215 08/21/97	ZAP Express	253	834	1	834	82	82	555	1	40	41	251
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	107	2327	1528	2327	465	465	409	1	30	31	284

	Gene No.	100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	257	689	218	655	314	314	559	1	18	19	95
	cDNA Clone ID	100	HEONN58	209119 06/12/97	pSport1	258	2377	5	2377	25	25	560	1	28	29	54
	ATCC Deposit No.: Z and Date	101	HCRAM28	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	111	1117	1	1117	1	413	1	19	20	21	
		101	HIBEK16	209627 02/12/98	Other	259	1193	69	1135	242	242	561	1	24	25	108
		102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	112	1313	128	1313	271	271	414	1	30	31	51
		102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	260	1262	26	1262	35	35	562	1	35	36	50
		103	HEBD182	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	113	1654	-	553	1654	709	415	1	-	-	32

[illegible]

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	114	1171	540	1171	337	337	416	1	30	31	163
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	261	1179	626	1161	335	335	563	1	30	31	253
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	262	1162	629	1131	942	942	564	1			18
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	115	842	373	800	100	100	417	1	65	66	174
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	263	735	290	735			565	1			
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	264	783	416	783		413	566	1	33	34	73

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	116	1640	187	1470	581	581	418	1	30	31	50
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	265	1638	301	1405	119	119	567	1	30	31	263
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	266	1455	148	1188	438	438	568	1	24	25	70
107	HE6DK18	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	117	952	418	906	499	499	419	1	28	29	120
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	118	1256	21	1079	301	301	420	1	30	31	159
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	267	1086	25	1050	227	227	569	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	Seq. ID	Seq. NT	Seq. Total	Seq. Clone	Seq. of 3' NT	Seq. of 5' NT	Start of 5' NT Codon	First AA of Signal	Seq. NO: Y	First AA of ID	Seq. AA	First AA	Secreted of AA	Last AA	ORF
109	HIPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	119	1143	171	1051	175	175	175	421	1	50	51	154			
109	HIPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	268	1003	21	1003	115	115	115	570	1	34	35	104			
109	HIPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	269	1234	174	1015	232	232	232	571	1	27	28	132			
110	HSXBL78	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	120	1782	1	1720	138	138	138	422	1	32	33	204			
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	121	610	18	609	50	50	50	423	1	30	31	67			
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	270	574	1	566	337	337	337	572	1	27	28	32			

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	Seq. ID	Seq. NT	Seq. Total	Seq. Clone	Seq. of 3' NT	Seq. of 5' NT	Start of 5' NT Codon	First AA of Signal	Seq. NO: Y	First AA of ID	Seq. AA	First AA	Secreted of AA	Last AA	ORF
112	HOEAP41	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	122	526	185	375	143	143	143	424	1	21	22	25			
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	123	2081	1179	1976	48	48	48	425	1	30	31	299			
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	271	1731	889	1626	886	886	886	573	1	18	19	28			
114	HTXGS75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	124	1717	764	1640	76	76	76	426	1		13				
115	HOVBA03	97902 02/26/97 209048 05/15/97	psport1	125	804	1	804	145	145	145	427	1	15	16	198			
115	HOVBA03	97902 02/26/97 209048 05/15/97	psport1	272	1320	77	637	280	280	280	574	1	22	23	40			

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	126	431	1	431	73	73	428	1	38	39	47
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	273	515	1	515	43	43	575	1	20	21	30
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	127	3752	3465	3752	748	748	429	1	30	31	370
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	274	2995	2738	2995	2777	2777	576	1	18	19	29
118	HASAS24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	128	1144	669	1144	896	896	430	1			30
119	HSIDN55	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	129	1830	1234	1830	1265	1265	431	1			24

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
120	HGBGZ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	130	1864	1505	1741	1578	1578	432	1	37	38	53
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	131	2041	1	1214	46	46	433	1	35	36	176
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	275	1990	8	1128	71	71	577	1	16	17	92
122	HOECP43	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	132	2012	853	1986	1127	1127	434	1	22	23	77
123	H2CBV31	97902 02/26/97 209048 05/15/97	pBluescript SK-	133	1669	670	1632	962	962	435	1	25	26	32
124	HPCAD23	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	134	1565	281	1565	274	274	436	1	25	26	30





Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	145	1021	526	1021	74	74	447	1	30	31	278
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	278	961	524	961	545	545	580	1	23	24	110
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	146	1285	5	1285	116	116	448	1	30	31	199
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	279	1228	9	1228	324	324	581	1	26	27	30
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	147	1386	169	1272	165	165	449	1	30	31	258
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	280	1327	169	1208	160	160	582	1	23	24	71

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
138	HMWGI25	97902 02/26/97 209048 05/15/97	Uni-Zap XR	148	2098	721	2044	784	784	450	1	18	19	87
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	149	1847	1689	1847	241	241	451	1	33	34	315
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	281	799	1	799		243	583	1	12	13	47
140	HMSKE75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	150	1569	113	1517	417	417	452	1	21	22	52
141	HCM SH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	151	1540	538	1540	48	48	453	1	30	31	383
141	HCM SH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	282	2196	270	2196	294	294	584	1	32	33	39

Gene No.	cDNA Clone ID	ATCC Deposit No. Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT Clone Seq.	3' NT Clone Seq.	5' NT Start of Codon	5' NT Signal Pep	AA of First SEQ NO: Y	First AA of Last Predicted	Secreted of AA	Secreted of AA	ORF
142	HTWCB92	97902 02/26/97 209048 05/15/97	pSport1	152	1719	690	1575	6	6	454	1	52	53	186
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	153	863	1	863	195	195	455	1	26	27	163
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	283	1185	277	1166	621	621	585	1			19
144	HFA MG13	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	154	1101	1	512	40	40	456	1	21	22	46
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	155	2031	669	2031	411	411	457	1	23	24	105
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	284	1634	615	1485	878	878	586	1	20	21	23

Gene No.	cDNA Clone ID	ATCC Deposit No. Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT Clone Seq.	3' NT Clone Seq.	5' NT Start of Codon	5' NT Signal Pep	AA of First SEQ NO: Y	First AA of Last Predicted	Secreted of AA	Secreted of AA	ORF
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	156	1981	1458	1809	1592	1592	458	1	23	24	70
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	285	1795	1458	1749	1562	1562	587	1	33	34	69
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	157	915	45	912	22	22	459	1	22	23	155
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	286	858	46	858	224	224	588	1	30	31	77
147	HSJAP03	209139 07/03/97	Uni-ZAP XR	287	915	1	915	22	22	589	1	22	23	155
148	HSKGO26	97903 02/26/97 209049 05/15/97	pBluescript	158	2117	51	1422	32	32	460	1	23	24	332
149	HCOAV96	97903 02/26/97 209049 05/15/97	Lambda ZAP II	159	2395	1509	2382	1440	1440	461	1			5

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
150	HSHCC16	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	160	2120	1223	2108	1416	1416	462	1			14
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	161	900	482	900	46	46	463	1	30	31	285
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	288	1517	783	1517	1062	1062	590	1			24
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	162	1003	1	1003	288	288	464	1	30	31	80
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	289	3865	217	1195	281	281	591	1	16	17	38
153	HTSFQ12	97903 02/26/97 209049 05/15/97	pBluescript	163	2196	1607	2180	1611	1611	465	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	164	1945	271	1840	299	299	466	1	63	64	96
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	290	1910	279	1818	355	355	592	1	39	40	69
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	165	2933	489	2871	258	258	467	1	30	31	399
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	291	3276	486	2838		525	593	1	45	46	308
156	HJPCJ76	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	166	2243	343	2221		341	468	1			1
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	167	1816	1130	1816	284	284	469	1	31	32	273



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	175	991	374	970	60	60	477	1	24	25	178
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	297	2416	1387	2413	1473	1473	599	1	18	19	25
166	HFKFX55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	176	1290	499	1290		688	478	1	25	26	52
167	H2LAO11	97903 02/26/97 209049 05/15/97	pBluescript SK-	177	2290	1	2290	173	173	479	1	22	23	62
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	178	549	1	549	11	11	480	1	21	22	27
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	298	545	1	545	17	17	600	1	21	22	27

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	179	1509	294	1352	92	92	481	1	30	31	339
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	299	1530	385	1530	562	562	601	1	23	24	61
170	HCFAE79	97904 02/26/97 209050 05/15/97	pSport1	180	1316	985	1250	995	995	482	1	26	27	32
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	181	777	1	777	51	51	483	1	30	31	48
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	300	997	244	997	300	300	602	1	23	24	29
172	HODCW06	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	182	791	1	791	14	14	484	1	29	30	38





Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	190	1014	703	1014	360	360	492	1	30	31	159
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	306	577	1	577		175	608	1			6
181	HAFUA18	97904 02/26/97 209050 05/15/97	pBluescript SK-	191	2779	2207	2630	1153	1153	493	1	30	31	279
181	HAFUA18	97904 02/26/97 209050 05/15/97	pBluescript SK-	307	2860	163	2860	21	21	609	1	30	31	232
181	HAFUA18	97904 02/26/97 209050 05/15/97	pBluescript SK-	308	876	275	876	302	302	610	1	32	33	34
182	HETBY74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	192	1923	30	1923	45	45	494	1	33	34	193

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	193	2346	1160	2286	178	178	495	1	30	31	205
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	309	2025	840	2025	971	971	611	1	18	19	21
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	194	3054	2004	3054	434	434	496	1	11	12	147
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	310	3026	1966	3026		2131	612	1			9
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	195	907	152	907	297	297	497	1	30	31	64
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	311	712	67	712	107	107	613	1	18	19	29



Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide

sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO: X and the predicted translated amino acid sequence identified as SEQ ID NO: Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO: X, SEQ ID NO: Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeech, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeech and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO: Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

### 10 Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

"Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, (1988); BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, (1993); COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, (1994);

SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, (1987); and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, (1991).) While there exists a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans. (Carillo, H., and Lipton, D., SIAM J Applied Math 48:1073 (1988).) Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in "Guide to Huge Computers," Martin J. Bishop, ed., Academic Press, San Diego, (1994), and Carillo, H., and Lipton, D., SIAM J Applied Math 48:1073 (1988).

30 Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux, J., et al., Nucleic Acids Research (1984) 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Aischul, S.F. et al., J. Molec. Biol. 215:403 (1990)), Besfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711 (using the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981)).

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set so that the percentage of identity is calculated over the full length of the reference polynucleotide and that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brulag et al. (Comp. App. Biosci. 6:237-245 (1990).) The term "sequence" includes nucleotide and amino acid sequences. In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter.

As an illustration, a polynucleotide having a nucleotide sequence of at least 95% "identity" to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone, means that the polynucleotide is identical to a sequence contained in SEQ ID NO:X or the cDNA except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the total length (not just within a given 100 nucleotide stretch). In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to SEQ ID NO:X or the deposited clone, up to 5% of the nucleotides in the sequence contained in SEQ ID NO:X or the cDNA can be deleted, inserted, or substituted with other nucleotides. These changes may occur anywhere throughout the polynucleotide.

Further embodiments of the present invention include polynucleotides having at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity

will encode a polypeptide identical to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence. Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., *J. Biotechnology* 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem.* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make

phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln; replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### **Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, and 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, and 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about"

includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.



Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In

preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the claimed invention.

### Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and lac promoters, the SV40 early and late promoters and promoters of retroviral LTRs; to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS,

293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46a, available from Stratagene Cloning Systems, Inc.; and puc99a, pKK223-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pVLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

### Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycle. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage

analysis establishes coinherence between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinherence is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the

present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erllich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20

millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention could be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

### Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules

may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

### Immune Activity

5 A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders 10 may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

15 A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: 20 blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, DiGeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome,

25 lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria. Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a 30 polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks 35 (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from

inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders. 5

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, 10 Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease. 15

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

20 A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD. 25

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g. septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.) 35

### Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thyroid, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellacea Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis,

Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### **Regeneration**

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

### **Chemotaxis**

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat



disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

### Binding Activity

5 A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

10 Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

15 Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

20 The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

25 Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

30 Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The

antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

5 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

### Other Activities

10 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

15 A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

20 A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, cardiac rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

25 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.



### Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone. A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining

whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the

amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a

polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an

amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least

199

90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated

200

polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

Vector Used to Construct Library	Corresponding Deposited Plasmid
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lambid BA	plasmid BA
pSport1	pSport1
PCMVVSport 2.0	PCMVVSport 2.0
PCMVVSport 3.0	PCMVVSport 3.0
PCR <sup>®</sup> 2.1	PCR <sup>®</sup> 2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128,256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988), Ailing-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Ailing-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation

of the fl origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the fl ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR<sup>®</sup>2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., BioTechnology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.

The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection

agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 µg of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl<sub>2</sub>, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to

remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

#### **Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO.X, according to the method described in Example 1. (See also, Sambrook.)

#### **Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprimer™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPD-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

#### **Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of

conditions : 30 seconds, 95°C, 1 minute, 56°C, 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

#### **Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9, (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis. Clones containing the desired constructs are grown overnight (ON) in liquid culture in LB media supplemented with both Amp (100 µg/ml) and Kan (25 µg/ml). The ON culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sub>600</sub>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic

agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number XXXXXX.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

#### Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a



stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Comassie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

## 20 Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Auographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAC373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as

required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five  $\mu$ g of a plasmid containing the polynucleotide is co-transfected with 1.0  $\mu$ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One  $\mu$ g of BaculoGold™ virus DNA and 5  $\mu$ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50  $\mu$ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10  $\mu$ l Lipofectin plus 90  $\mu$ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to S9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.



After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200  $\mu$ l of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5  $\mu$ Ci of  $^{35}$ S-methionine and 5  $\mu$ Ci  $^{35}$ S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### 25 Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),

pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., BioTechnology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the

naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100-200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

#### Example 9. Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Trautnecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having

more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No.209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGGCCAAATCTTCTGACAAAACATCACATGCCACCGTGCC  
CAGCACCTGAATTGAGGGGTGACCCGTCAAGTCTTCCCTTCCCCCAAACC  
CAAGGACACCCCTCATGATCTCCCGAGCTCTGAGGTACATGCGTGTGGT  
GGACGTAAGCCACGAAAGACCCTGAGGTCAAGTTCAACTGTGTACGTGACG  
GCGTGAGAGGTGCATTAATGCCAAGACAAAGCCGCGGAGGAGACGATACAAC  
AGCAAGTACCGTGTGTGACGCGTCTACCGTCTGCACAGAGACTGGCTG  
AATGGCAAGAGTACAAAGTGCAGAGGTCTCCACAAAGCCCTCCCAACCCCC  
ATCGAGAAAAACCATCTCCAAAGGCCAAAGGGACGCCCCGAGAAACACAGGT  
GTACACCCCTGCCCATCCCGGGATGAGTGAACCAAGAACACGAGTCAAGCT  
GACCTGCGCTGTCAAAAGGCTTCTATCCAAAGCGACATGCGCGTGAAGTGGGA  
GAGCAATGGGACGCGGAGAAACAACACTACAAGAACACAGCGCTCCCGTCTGG  
ACTCCGACGGCTCTTCTTCTCTCTACAGCAAGCTCACCGTGGACAAAGAGCA  
GGTGGCAGCAGGGGAACGCTTCTTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGACGAGAGAGCTCTCCCTGTCTCCGGGTAATGAGTGC  
GACGGCCGGAGCTTAGAGGAT (SEQ ID NO:1)

**Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody

whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

**Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhitaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10<sup>5</sup> cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhitaker))/10% heat inactivated FBS(14-503F Biowhitaker)/1x Penstrep(17-602E Biowhitaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstep, or CHO-5 media (see below) with 2mm glutamine and 1x penstep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

*HGS-CHO-5 medium formulation:*

**Inorganic Salts**

CaCl <sub>2</sub> (anhyd)	116.6 mg/L
CuSO <sub>4</sub> ·5H <sub>2</sub> O	0.00130
Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O	0.050
FeSO <sub>4</sub> ·7H <sub>2</sub> O	0.417
KCl	311.80
MgCl <sub>2</sub>	28.64
MgSO <sub>4</sub>	48.84
NaCl	6995.50
NaHCO <sub>3</sub>	2400.0
NaH <sub>2</sub> PO <sub>4</sub> ·H <sub>2</sub> O	62.50
Na <sub>2</sub> HPO <sub>4</sub>	71.02
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	.4320

**Lipids**

Arachidonic Acid	.002 mg/L
Cholesterol	1.022
DL-alpha-Tocopherol-Acetate	.070
Linoleic Acid	0.0520
Linolenic Acid	0.010
Myristic Acid	0.010
Oleic Acid	0.010
Palmitic Acid	0.010
Palmitic Acid	0.010
Pluronic F-68	100
Stearic Acid	0.010
Tween 80	2.20

**Carbon Source**

D-Glucose	4551 mg/L
-----------	-----------

**Amino Acids**

L-Alanine	130.85 mg/ml
L-Arginine-HCL	147.50
L-Asparagine-H <sub>2</sub> O	7.50
L-Aspartic Acid	6.65
L-Cystine-2HCL-H <sub>2</sub> O	29.56
L-Cystine-2HCL	31.29
L-Glutamic Acid	7.35
L-Glutamine	365.0
Glycine	18.75
L-Histidine-HCL	52.48

H <sub>2</sub> O	
L-Isoleucine	106.97
L-Leucine	111.45
L-Lysine HCL	163.75
L-Methionine	32.34
L-Phenylalanine	68.48
L-Proline	40.0
L-Serine	26.25
L-Threonine	101.05
L-Tryptophan	19.22
L-Tyrosine-2Na-2H <sub>2</sub> O	91.79
L-Valine	99.65

#### Vitamins

Biotin	0.0035 mg/L
D-Ca Pantothenate	3.24
Choline Chloride	11.78
Folic Acid	4.65
L-Inositol	15.60
Niacinamide	3.02
Pyridoxal HCL	3.00
Pyridoxine HCL	0.031
Riboflavin	0.319
Thiamine HCL	3.17
Thymidine	0.365
Vitamin B <sub>12</sub>	0.680

#### Other Components

HEPES Buffer	25 mM
Na Hypoxanthine	2.39 mg/L
Lipoic Acid	0.105
Sodium Putrescine-2HCL	0.081
Sodium Pyruvate	55.0
Sodium Selenite	0.0067
Ethanolamine	20uM
Ferric Citrate	0.122
Methyl-B-Cyclodextrin complexed with Linoleic Acid	41.70
Methyl-B-Cyclodextrin complexed with Oleic Acid	33.33
Methyl-B-Cyclodextrin complexed with Retinal Acetate	10

Adjust osmolality to 327 mOsm

#### Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	ISRE Ligand	tyk2	JAKs Jak1 Jak2 Jak3	STATs	GAS(element) or
5	IFN family IFN- $\alpha$ /B IFN- $\gamma$ (IRF1>Ly6s>IFP) IL-10	+	+	-	1,2,3 ISRE GAS
10	gpl30 family IL-6 (Pleiotrohic) (IRF1>Ly6s>IFP) IL-11 (Pleiotrohic) OnM (Pleiotrohic) LIF (Pleiotrohic) CNTF (Pleiotrohic) G-CSF (Pleiotrohic) IL-12 (Pleiotrohic)	+	+	?	1,3 GAS
15	IL-2 (lymphocytes) IL-4 (lymph/myeloid) >Ly6(lgH) IL-7 (lymphocytes) IL-9 (lymphocytes) IL-13 (lymphocytes) IL-15	-	+	-	1,3,5 GAS GAS (IRF1 = IFP)
20	gC family IL-2 (lymphocytes) IL-4 (lymph/myeloid) >Ly6(lgH) IL-7 (lymphocytes) IL-9 (lymphocytes) IL-13 (lymphocytes) IL-15	-	+	-	6 GAS GAS (IRF1 = IFP)
25	gpl40 family IL-3 (myeloid) (IRF1>IFP>Ly6) IL-5 (myeloid) GM-CSF (myeloid)	-	+	+	5 GAS
30	IL-3 (myeloid) (IRF1>IFP>Ly6) IL-5 (myeloid) GM-CSF (myeloid)	-	+	+	5 GAS
35	Growth hormone family GH PRL EPO CAS>IRF1=IFP>Ly6)	-	+	+	5 GAS(B-
40	Receptor Tyrosine Kinases EGF	+	+	-	1,3 GAS (IRF1)
45	PDGF CSF-1	+	+	-	1,3 GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994)), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5'-GGCGCTCGAGATTTCGCCGAAATCTAGATTTCGCCGAAATGATTTCGCCG  
AAATGATTTCGCCGAAATATCTGCCATCTCAATTAG.3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5'-GGCGCAAGCTTTTGCAAAAGCCTAGGC.3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5'-CTCGAGATTTCGCCGAAATCTAGATTTCGCCGAAATGATTTCGCCGAAATG  
ATTTCGCCGAAATATCTGCCATCTCAATTAGTCAGACAACCATAGTCCGCC  
CTAACTCCGCCCATCCGCCCTAACTCCGCCAGTTCGCCCATTTCTCCGC  
CCCATGCGCTGACTAATTTTATTATTATGACAGAGCGCCGAGCCGCTCGGC  
CTCTGAGCTATTCCAGAGTAGTGAGGAGGCTTTTGGAGGCGCTAGGCTTT  
TGCAAAAAGCCT.3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS-SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a

neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NF- $\kappa$ B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF- $\kappa$ B/EGR, GAS/NF- $\kappa$ B, IL-2/NFAT, or NF- $\kappa$ B/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

### Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jak-STAT signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/Neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200  $\mu$ l of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI

+ 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies) with 10  $\mu$ g of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50  $\mu$ l of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1 ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat/GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200  $\mu$ l of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50  $\mu$ l of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35  $\mu$ l samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at 20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 40°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

#### **Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in

5 Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Khatabanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^6$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

25 These cells are tested by harvesting  $1 \times 10^6$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

#### **Example 15: High-Throughput Screening Assay Identifying Neuronal Activity**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

10 Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

15 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1) (Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

20 5' GCGCTCGAAGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)  
5' GCGAAGCTTCGGCGACTCCCGGATCCGCCCTC-3' (SEQ ID NO:7)

25 Using the GAS/SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS/SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

30 To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

35 PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.



Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I- $\kappa$ B is phosphorylated and degraded, causing NF- $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:

5'-GCGGCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC  
TTTCCATCCTGCCATCTCAATTAG-3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5'-GCGGCAAGCTTTTCCAAAGCTAGGC-3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5'-CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGGACTTTCCGGGACTTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCAATAGTCCCGCCCTAAGCTCGCCCA  
TCCCGCCCTAACTCCGCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACT  
AATTTTATTATGACAGAGGCCGAGGCCGCTCGGCTCTGAGCTATTC  
CAGAAAGTAGTGAGGAGGCTTTTGGAGGGCTTAGGCTTTTGCAAAAAGCTT:  
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the

NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

- Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### Example 17: Assay for SEAP Activity

- 10 As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

- 15 Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

- 20 Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room

- temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

- 25 Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4

15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension.

The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular

signaling even which has resulted in an increase in the intracellular Ca<sup>++</sup> concentration.

**Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity**

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are

used to cover the Loprodyn Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

- To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyn plates (20,000/200µl/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 µl of the supernatant produced in Example 11, the medium was removed and 100 µl of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 40°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 µm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 40°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

- Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- The tyrosine kinase reaction is set up by adding the following components in order: First, add 10µl of 5µM Biotinylated Peptide, then 10µl ATP/Mg<sub>2</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10µl of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycero-phosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5µl of Sodium Vanadate(1mM), and then 5µl of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10µl of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 µl of 120mM EDTA and place the reactions on ice.

- Tyrosine kinase activity is determined by transferring 50 µl aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This

allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300µl/well of PBS four times. Next add 75 µl of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5µl/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

- Next add 100µl of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

#### **Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

- As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

- Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1µg/ml) for 2 hr at room temp. (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

- A431 cells are seeded at 20,000/well in a 96-well Loprodyn filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 µl of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

10

**Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products is then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals is identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenin deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

35

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

10

**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect soluble polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on

35

the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

### Example 23: Formulating a Polypeptide

5 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

10 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1  $\mu\text{g/kg/day}$  to 10  $\text{mg/kg/day}$  of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01  $\text{mg/kg/day}$ , and most preferably for humans between about 0.01 and 1  $\text{mg/kg/day}$  for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50  $\mu\text{g/kg/hour}$ , either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

20 Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

30 The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919; EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Siddman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric

acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

10 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

15 Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

20 The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as poly(vinylpyrrolidone); amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium, and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

30 The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of

about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile.

5 Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

#### Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

30 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

#### Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

5 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

#### Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

25 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to

transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphiphotropic PA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is being produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

(1) GENERAL INFORMATION:

(i) APPLICANT: Human Genome Sciences, Inc. et al.

(ii) TITLE OF INVENTION: 186 Human Secreted Proteins

(iii) NUMBER OF SEQUENCES: 644

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

(D) STATE: Maryland

(E) COUNTRY: USA

(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: March 6, 1998

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:



## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: A. Anders Brookes, Esq.  
(B) REGISTRATION NUMBER: 36.373  
(C) REFERENCE/DOCKET NUMBER: PS002.PCT

5

10

## (vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504  
(B) TELEFAX: (301) 309-8439

15

20

## (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 733 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GGCAGATCT TCTGACAAA CTCACATG CCCACCTGC CACGACCTG 60  
AATTGAGGG TGACCGTCA GTCTTCTCT TCCCCCMAA ACCGAGGAC ACCCTCATGA 120  
TCTCCCGAC TCTGAGTGC ACATGCTGG TGTGTAGCT AGCCACGAA GACCTGAGG 180  
TCAGTTTCA CTGTAGCTG GACCGGTGG AGGTGCATAA TCCACAGCA AAGCCCGGG 240  
AGGACGCTA CACAGCAGG TACGCTGTGG TCAGCTTCT CACCTTCTG CACGAGACT 300  
GGCTGAATG CAGGAGTAC AAGTGCAGG TTCTCAGCAA AGCCCTCCA ACCCTCATGG 360  
AGAAACCTT CTCGAAAGC AAGGCGAGC CCGAGAGCC ACAGGTATC ACCCTGCCC 420  
CATCCCGGA TGACCTGACC AAGAACGAG TCAGCTTGAC CTCCTGTGC AAGGCTTCT 480  
ATCCAGCGA CATTGCGGTG GAGTGGGACA GCAATGGCA GCGGAGAGC AACTACAGA 540  
CCAGGCTCC CGTCTGTGAC TCCAGCGCT CTCTTCTCT CTACAGCAG CTCACGTTG 600  
ACAGAGCAG GTGCGAGCAG GCGAGCTCT TCTCATGCTC CTGTATGAT GAGGCTCTGC 660  
ACACCACTA CAGCGAGAG AGCCTTCTCC TGTCTCCGG TAAATGATG CGAGGCGGC 720  
GACTCTAGAG GAT 733

55

## (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser  
1 5

10

15

## (2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 86 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGGCTCGAG ATTTCGCGA AATCTAGATT TCCCGAAT GATTTCGCG AATGATTC 60  
CCGAAATAT CTCGATCTC AATTAG 86

30

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGGCAGCT TTTTCGAAG CCTAGGC

45

## (2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 271 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATT CCGCGAATC TAGATTCCC GGAATGATT TCCCGAAT GATTTCGCG 60

60

AAATATCTGC CATCTCAATT AOTCAAGAC CATTACTCCG CCCCCTACTC CCCCCTATCC 120  
GCCCCCTACT CCCCCCAATT CCCCCCAATT TCCCCCAAT CCCCCCAAT TTTTTTTTAT 180  
TTATTCAGAG GCGGAGGCGG CCGGAGGCTC TGAGCTATTC CAGAGTATAT GAGGAGGCTT 240  
TTTTGAGAGC CTAGCTTTT GCAAAAAGCT T 271

(2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCCCTCAGG GATGAGAGG ATAGACCCG GG 32

(2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCGAAGCTTC GCGATCCCG GATCCCGCT C 31

(2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGAAGCTTC CC 12

(2) INFORMATION FOR SEQ ID NO: 9:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCGGCGTCA GCGAAGCTTC CCGGAGACTT TCCGGGAGCT TTCCGGAGCT TTCCATCTCG 60  
CAATCAAT TAG 73

(2) INFORMATION FOR SEQ ID NO: 10:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTGAGAGGGA GTTCCCGGCG GACTTCCCG GAACTTCCG GAACTTCCA TTCCATCT 60  
CAATTATCA GCAACCTTAG TCCCGGCTT AACTCCGCG ATCCCGGCG TAACTCCGCG 120  
CAATCCCGCG CATTCCCGCG CCGAGCGCTG ACTATTTT TTATTTTAGG CAGAGCGGGA 180  
GCGGCGTCC GCTCTGAGC TATTCAGAA GTAGTAGAGA GCTTTTTCG GAGCGCTTAG 240  
CTTTGAAA AAGCTT 256

(2) INFORMATION FOR SEQ ID NO: 11:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 582 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GCGAGAGGT AATTCCTAC ACAATTTCC AAGCATTTT GTAGCTGAA AAAATGCA 60  
GCAAGCTTT AAGACTCTG GATCCATTA TAGATATGT ATAGCTGAA TGTGTATTC 120  
AATGCTTTT TCTCTTTT CCTGAGCA AAAATTTT TTATTTTCA TCCAAATGT 180  
TTTATCTT TGTATTTT TTAAAAATCC TTTCTCTT ATCATCTCT TTTTCTGT 240  
TTTAAATGCA CTGACTTCA CTTGAAAGT GATTTCTCC TTGTATCTT ACAATATGT 300  
GATATGTA TTTCAGAAC AGATCTAGT TTGAGGAA GATATGTA TTTGTTTATA 360  
ACAAAAAC TGTCTCAT TCTGTGAAT TGTCTTTA AATTTCTT TTACAGTGT 420

AGCCNACTG AGATACCTG ATGGCTTGA TTCTTTTCA TGTGCTTAC CATCTATTT 480  
AGCCACTGAG CCTTTTATTA TTCTCTATT TGTAAAGTTT ATTGTCTTA ACTCATTTA 540  
TAATATTA CTATTACTGTTTATCTGT TTCTGAAAAA AAAAAAAAAA AA 582

10 (2) INFORMATION FOR SEQ ID NO: 12:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 465 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
GTTTGGGGT GAGGCCGAGC TGCTGGGGG CTTCGTGCG GGCAGGACA CAGCTACTG 60  
CAGGCGGGG GCGCTTGCT ATGATGTC TCACCCAGGG GGGGCTCTG CCTCTACTC 120  
GTGCGAGCC CACTTGCCAG GAGAGGCC TCCCAGCC TTGAGGGCTG CTGCGAGTCA 180  
CCTGTGGAA TGGACTAAA GAGCCTTGT GTGGAGAG GTGCTGCCA AACAGCTGC 240  
TCTGTGCTC CAGCCAGGCC CTCTGAGG GAGGGGAG GACTCATAG GACCTGCTG 300  
GACCCCTGCA GGGCAGGCG CTGGGGCG AGCCAGCA TTGCGCTG CTGCCCCAA 360  
GGGACAGGA AGCCTTGGG GCTCTGCC TTCTGGACA AGGCCCCCTG CCTTTGCTC 420  
ACATAACTG TACATATT TCAATAAG CCTTTTAT AAAA 465

40 (2) INFORMATION FOR SEQ ID NO: 13:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 474 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
ATGCATTC TCCTCAGC CTTCCTGT GTGCCACTC TGCTCTTG TGTGTGCC 60  
ATATGCTAG GCTTCTGCC CTCTAGAG GGTCTCTTGA TAGATTAGA AATAAGATG 120  
AGTACATTT CTAATGTCA TATAGAGG AGCCAGCA CATGCTTTT AATTAAGG 180  
ACAGTGTCA TCTTTTAC TGGCAATAG AACCTTGGT TCATCTGCT GAGCTAGGC 240  
CTTTAACA GCTTCTGT TTCTATTT TCTAGTGT TTGCGAGGT TTATGCGAA 300  
AGATATGTT CGTTTAAA TATTTCTTA TTAGGCGGG GGTGTGCT CAGCCTGTA 360

ACCTAGCA TTGGGGCTG AGCGGTGGA TCAGGAGTC AGGAGATGCA GACCATCTG 420  
GSTAACATG TGAACCCCG TCTCTACTA AATACAAA AAAAAAAAAA AAA 474

10 (2) INFORMATION FOR SEQ ID NO: 14:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 314 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
TTATGTGG GAGCAGACC TGATAGCCAG CCTTTACATG GGAGTATAT TCTGTCTCC 60  
ATCTCATAG CCCAGTACC TGAGCGAGA TGTATTATC CAACCAACT GTCTCTTAT 120  
CATGATGG TTTCAGATA GTTATTTTC ATCATTTCCA TTGTAGCTC TACAGTGTT 180  
TATATAATT TCTATCTTT TAAGTCTCT CCTCAGTCC TGTGTATAT AAACATCAT 240  
CTCTCAGG CAGTTGAGT CTGCATCTC CTTATGGGG CAGAGCTGT TTGAGAGAG 300  
AGATHTAC TTCC 314

30 (2) INFORMATION FOR SEQ ID NO: 15:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 613 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:  
CTCATATTC CTTCTGCTA AAGTGAACA TGCATGAT CATCTGCT TTATTATAT 60  
ATCTTCTTA TGTGGCAG CAGACAGA AGTAGAAGA AGATGTGT AGCTCAGA 120  
ACCACTAA TCTATCTAT GCGTGGGT CACCCAGCT GCTTGTGGA TTGTGTCTA 180  
CTATAACGA GCTGCCAGG AGACTGAGA GTACGCTCC TTAAGCACT TAACTAAGC 240  
CTACTCTTC CTTTCCACC AACATGTTC CGACTCAT CTCCTTCCR AGTCCCTTT 300  
TCTGCCCG ATGCAATTC CATTTAATA ATCTCAAGT GAATGTCCA CACAGATTC 360  
CATTTAATT AGCATACCAT AGTTTTTGT CAATTTGCT TTCAGAGAC TCCATTGCA 420  
GCTGCTAGA CAGCTAAG GCGGGGCTC TTGAGCTTT CCCGATAGT TTCAGCTGCA 480  
ATAGTCTTA GCGAGATGC CATGAGCTC CTGCCCCACT GTATTACTGG GGAACCTGG 540

ATTGGCTAGA AGTTGATCTT CCTGTAACCT TTCTGAGCTC TTACATTTA CTCGTGAAC 600  
CCAAATATCC CAC 613

(2) INFORMATION FOR SEQ ID NO: 16:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 356 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CCCCCCCC TGAACCTGG GCTGTGAAG TTTTGGCTG TGTGTGCTT TGTGTGTGC 60  
GCTGTGTGTG TGTGTGTGCA CTCCTGTTC AAGTGTGAG CAGCCATCA TGAAGCCCC 120  
TTATTTTATG TTGCATGCA CCAAGTCTCC CCCCCAGC CTCCTGTGAG TCCCTCATG 180  
GTATGTGTGC TCCCTGTGCA AGAAGCTGCA TTGTGTGAAA ATTGCATCAT GTATTATAT 240  
GGAAAGCAT TTGTGTGTG TTATGTTTAT TACATATCA TTGCATGTG GAACTGCAT 300  
CCTGTGAAC GTATTATGTC GTTTTACCA AGAAGCAGAG GGGAAAAA AAAA 356

(2) INFORMATION FOR SEQ ID NO: 17:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 414 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAAATTAAT CCCCCGCTT TTAAAGCTA CTCGTGAAT AAGTATGAG TATCTACTA 60  
GGAGAGTGG AGTGCACCA ACAAAGCTA TAAAGGAAA TGATGTACTT TTTAAAAAT 120  
ATTGCATTA ACTGTCTAG GATACTCTC TTGAGGCTT GAAATCTC TTCTGTGAA 180  
TTTGCATATC CACTGCATG CTGTACCAA AGATTTAAT CTCATGATG CATTTCTC 240  
TCTCCAGAA AAAATGACTA CAGAGCTC AAGAGTATG CTTGTGTGT GAGGAGTTA 300  
CACTATGTT TTCTCTGT TCAATATGT ATTACAGAA GACTGTCA TCAAGAGAG 360  
TACTAGACTA TCTTATAGC TTGTGATTTG ATCAAGCTT TAAAAAAA AAAA 414

60

(2) INFORMATION FOR SEQ ID NO: 18:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 469 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AATCAACAT GCATACAAA TGATCTGCT GTTAATGTT GACTGTGAC CCAATGTCT 60  
CAGAGACTC GAGTGTAGG TGTGTCTTC CTTGTGTAC ATTAAGTGGAA TAAATCAACC 120  
CTTCATATTC TTCAAGCTTA GGAAGTCTG GCAGAGCTG CCCCATTT TGTTCGGGT 180  
GTTCGGCTTG CTTCACCGGG GATGTAGGCT ACTGTGCA GAGACGAAG GGGTGGCTT 240  
GCCAGAAC ATGAGGAGG CCAAGACTC TGGAGGAAA GCAGGCCA AAGAAAAAC 300  
GATTACTT AAGTCCAAA CTCAGAAC CTCGGCAC TGAAGAAAT GTTTTGGGC 360  
GATGTGTGT TGAAGAGAT AGATGAGAT TATCTCTGC AAAAACTCT AAGAGCTTC 420  
ACTGTCTGT ATTCTCTTC ATACTGCTT ACCCCCAAT TAAATTCAG 469

(2) INFORMATION FOR SEQ ID NO: 19:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 550 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CCCCCCCC CCCCAGACT TTCAAGATC ACCCCCCAG ATTGTGGGTT GGGTTGGCC 60  
TACTCAACC TGAAGCTGCC TGAAGAAC TCACTGCTT GCTCCATC CCCCCTACTG 120  
GTGCAGGAAT GCAGCTGGA GCTGGCTGCC TTGTGTGAG CTTGTGTG CCCCAGCCCC 180  
AAGCCCCAG CTCGTCTGA ATTGACATCA GTCTGTCTT GAACTGCTC CCCCAGCTT 240  
GGGATTAAT CCAAGAACT TTATGTTC TTAAAGCTA GCAGCTGTC GCATCTAGG 300  
ACTGTGAGG GTAGGCTGAG TTGCAGTCA GTCTTGAGG GTCTGTAGG ATCTGACTG 360  
AGAGCTTGC TACTCCCAA GCCAGGCC ATCAAGCAG CTCGTGTA GCCAGCTGCC 420  
TTGTAGTGC TGAAGTCA CAAAGCTG CAGCTGTC GCTCAATTA AAGGATTTG 480  
ATGAGCCGAT GGGCTGTGA GCGAGCCAT TAAAGATCT GGTGTGTTT TGGAAAAA 540  
AAAAAAA 550

60

## (2) INFORMATION FOR SEQ ID NO: 20:

5

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 741 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 TCTTGAAAGAG TGTACAGTAC AGGATATTA TATGTAAAGT TTAATTCAC AGGGTTTCT 60  
15 TGGCTCTGCA TATATTATTA GAAAGAGAG TTGTAAAGT GGCACAGTAT TCCAGATAC 120  
TTTTCAGTTG GGGCTTTCT TCTGGTTCTT TAATTTGAAA CTTAGATACA TCCAGTAAAA 180  
20 ACTAGAGAAA TGACTTTTAC CTTGGGAG AGCGAAGTTT TGTGTAAAA CTTATTTCT 240  
AGCATGCTT CAGCAAGTTG TCCAGACC TAGATTGTGA AGGACCACT GTTCTTCTGT 300  
TGTACAGCT CCGTGAACCA TTGTTCAGAG GACCAATGTC ACATGCTTTC ATGGGCAATG 360  
25 NCCATGGGAG CATTGCTGAG ATATCTCTCT ACAGTATTTG CTCTTCTGG AGGCTGATAC 420  
ACAGGCTTC TCTTCACAT GATCATTTGC AAACCTGCCC CAGCCCTTAC CATTCAATGT 480  
GGAGAGAAA CAAGACTCC CTGAAGAGA GTCCAAGCTA CAGATACACA CCGTGTGCAT 540  
30 TCGGCTGTC ACTTCTGCC TCCGACTTCT GTATCTTCAG AGATCTGCG TCGATTTTC 600  
CTTAACCTCA GCTGACTTCC CTGTGAATGT CTATGTCTAG TTCAGGCTT CCAAGCATTC 660  
35 ATTGTACAG TGGTAACTCC CAATGAGGCT TCTGTATACA TTGTGTGTC TTATTTCTGC 720  
ATTAAAGAA ATGATTTTCC C

40

## (2) INFORMATION FOR SEQ ID NO: 21:

45

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 991 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

5 GGCACAGTCC TCCCTCTGGG AGTTTTTCT TTTTCAGAG GAGAGAGGC TTTCAGAT 60  
15 ATGTGTGTA GAGTGTGGG CAGAAATCT GGGACACAC CACACCAAT CTCTCTTAA 120  
TCCAGCTAT TTGCTTTCTA TCCACACTAT GTTTCCAGTG TCTCTGGCT GTTTCCAGA 180  
55 GCAACAGAA ATGATTAAT CTCTGTGAG TTGTTTATTT GTTCTCACT TTGTTTACA 240  
60 CTGTATTTTC TGAATTAAG GGTCTCTGAG AATTAAAGG GAATAGTAA ATAAAGTAAA 300

5 ACTCAGCTTG AAGGAATAT ACTAAATTA GATAAAGCTG ACCTGTAGAT ATAGCAGCTT 360  
ATAAAGCTTA GAGTGTGCTA AGTTGAGTGC AAATTTTCTT CTGATCTTTC TGATCCGAA 420  
CAAAAAGCA GTCTGTTTG TTATGTGATT GGAATGGAC CCGAGAAGAG AGCATGCTGT 480  
GTCTTGTGG GACAGGAAG CTTCGTGCA CCAAGTCTGA ACCACCACT TCATGCTGAC 540  
10 ATGATTTATG TCGTGAACA TATTTACAC CGGCTTGCA GTAAACACTT GTATGTTGT 600  
CGATGGAAA CGGTCTCTT CCGCTAAGC AGGGCTGTT GTGCAAGGGA ATGCTCATC 660  
TCTGCTAAA ACACAGCTTC CATTGATG TATGCTGCTT ACTCAAGAG TGTGTCTCCA 720  
15 AACAGCTTT GGGAGTCTT CTTGATTTCA TGGATGAAC CTGCAACATC TTGAGCACTG 780  
AGTTAACAT AGTCTCTTAA ATAACTCTCC ACAAGTTTT CTATGTTTAT CTCTACATCC 840  
20 AGGCTGTGCA GCAAGCTGTT CAAGTCAATA TTTCTGGGA ATATTTTCCA GTGTTTATTT 900  
GCACTTTAGC CCACTCTGTG TAGCTTATTT TCTTCTAACC TCACCAATTA TCTGATTAAT 960  
25 ACTCAATTT AGGGGACTG TATTTGCTTT A

## (2) INFORMATION FOR SEQ ID NO: 22:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 653 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

5 CCACGCTCC GGAATTCGCC TGAGATCTT GGGCTATCTT TGACAGGGA TTCTTGCAAG 60  
15 TTGATGCTTT CTACAGTCA ATTAGTCTAG TCCCAAGA TGGAGAGCTT GAGTTCTCAC 120  
AGATTTGATG AAGATGGAGA AAACACACAG ATTGAGGATA CGGAACCCCT GTCTCCAGTT 180  
45 CTCATTTCTA AATTTGTTCC TCTCAAAAT GATAGTATCC TGATGAATCC AGCACAGAT 240  
GTCAGTAC ACTGAGTCA GAATGATGAC AAACAAGG CAGATGATAC AGACACAGG 300  
GATGACATTA GTATTTTAC CACTGTTTC AAGGCGAGAG AAGAAACGGT AGCAGAGAA 360  
50 GTTTGATTTG ATCTCACTTG TGAITCGGG AGTCAGGAG TTCCGTCAAC AGCTACTGCA 420  
TCTGAGGCAC TTTCATGTT GTTAGTACAG GAGGAAGCTA TGGAAATTA AGAAACCAT 480  
55 CCAGAGGAG GGTCTTCAGG GTCTGAGGTG GAAGAAATCC CTGACACACC TTGTGAAGT 540  
CAAGAGAG ACTCAAGA AGAAATATG GAGCTGTTT CATTGCACT TTCTCTGACT 600  
60 GAATCTAGT CCAAGGGTT GTGTCTTGG AGCATCCA AAAAAAAAA AAA 653

## (2) INFORMATION FOR SEQ ID NO: 23:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1486 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

5 GCGAGCTGCA CAGCTGCA GCGAGAGTGG CTGCCCCGTC GGTGCGGAGC GGTGGGAGAC 60  
15 CTGCGGAGAG AACCTGGCTG AACCCGAGAA CCCCAGGAGG CATTGGCTGC 120  
TTGGCTTGGG GAGCTTGAGA GCGCGAAGAC GCTTGTGAGA GCGAGAGAG TCTTTCAGC 180  
20 GCTGGGCACT GGTGCTGCA GAGACTGCA TTGAGCTCAT GGTGCTCAT GCGAGAGTGC 240  
TTGGTTTGGG GGGGTGTCGG GCTGTCAATG CCACTGGGCA CTTTTCAGAC ACAGTTTGGC 300  
25 TGAATCCCAT GAGATTCTTG AACATCCGCT ATGGTGAAGT GTTGCCGAGC ACCATGTGAG 360  
GAGAGATGCT TTGGCTTGAC ACTGAGATCA TGGGTGTCTG CTGACAGACC CTGCTGGTGG 420  
CGGTGGTGGC CCGAAGAGTG GAGTTTAAAG AACGAGAGAA GCAAGTGCAC AACCTTCATCA 480  
30 TGGATATCCA GTTAAACAAA GAGATTAAGG AGTTCGCTGC CCGAGTCTCA CAGAGAGCT 540  
GAGTTTCTTA CAAACATACT GCGAGAGAGG AGTTCATGAC TGCCTCGAGG CATTAGGCGA 600  
35 AACGATGAGA CTTCATGGTG GAGATTCCA AGATGCAAT GATTCCTGAT GAGCTGAGAC 720  
AGAACTGAGG CAGCTGCAC CCGGCCCCGG AGAAGACAGT TGAACAGCTG GCGGGAGAGC 780  
40 TGGATGCCCT GAGTGAAGTG GTTGAAGAGT CCGTGGGGGC GAGGCACTTT CAGAGAGCGA 840  
GCGAGAGATC CAGATAGCTG GAGCCAGAGG GAGGAGACAG GGTACTTTCC CAGATTACTGA 900  
45 GGTGGTGAAC ATGTTCTCTG CCACTCTTGA CCGAGCCCTG AACAGAGAC CTGAGATGCA 960  
AGAGAGAGAG GGGGCCCCGG CTGGAAGTGG GTTGGCTTGC TGAATGCTGC TGGAGGGAGC 1020  
GCTGGGAGAA GTGGGAGAGC GTTGGCTGAC CTGAGGGCCC AGGTGGGAGC ATGCTGACCC 1080  
50 CCACTGTCGA TACCTGATTC AAAAAGACTC TCAATATGCT GGTATTAAGG AGCTGAGCT 1140  
CTGAGTTTGA AGTGCAGGCG ACTGAGAGCA GAGTCTCTGG GTTCCCTGGA AAAGAGGTAC 1200  
TTGGGGCCCC GATTCAGAGT TGTGGAGAGC TTCAATTAAC GCTGGCGGAG CTGAGAGACT 1260  
55 GGTATTAAGG CTGGGGCGGG GCTGGAGTGG GCGGCCCCCTG GTGGAGACAC AAAAGAGACA 1320  
CGATTTTTCC AGAGCTGAG AGAGAGACTG GTGGGAGAGG AGAAGTTTAA CTGACAGACC 1380  
60 TGTGCTCTTA TCTTGTAAAT AAATGTAAAG GCGAGAGAAA AATTAAAAAA AAAAAAAAA 1440

## (2) INFORMATION FOR SEQ ID NO: 24:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2123 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5 AACTGAGGG GGGGCCCCAC CCAATGCCCC TATAGTAAAC GTAAATN 1486  
10 CTGCGGCTTT TCTCTGCGA GGGGAGGTCC CCGCTTCCCG TGGAGGCTCC GAGAGAGACC 60  
15 GTCGCGGCTT CTGCTTCGCG ATGCAATGTGC TGCCTGCGGC GCGCGGAGCG TTCCGCTGTC 120  
CTTTCAGCTT CTGCTTCGCG ATGCAATGTGC TGCCTGCGGC GCGCGGAGCG TTCCGCTGTC 180  
20 TTGGGTCTCT GCTTCCGAGG CCGGCTCCCG CCGAGCCAGC GAGCATGTGG GGGATCAGAA 240  
AGCAAGAGAC GAGAGAGCG GAGAAATCCA CCAATGATAT CTATGAGGCC CAGCATGTGCA 240  
25 GAGAGATTA GAGAGAGGCA GTGGTTGAGA CAGGGGTGGG GTTCCGAGGA TGTACCTGT 300  
GCTTAAAGAG TCTCTCTGCT GCTGGGAGAA AGACAGTAA GTTTGCTCT GAGAGAGTAC 360  
TTTGTGCGCA TGCATGCTCT GTTAAATCTCT GAGATGGGAG AATGTGCTTC ATGGCTTAA 420  
30 CAGATTCGCC CAGATGCTCT CATTGCCCCC AAGCATGAGA AGTCTGAC AGATTAATAC 480  
AGTTCCTGAC AGAAGAGCTA AGCTTTGCT CCGAGTTTTC TTCTGAGCT GCTCTTGAAT 540  
35 TACCTTTTCT AATTATTAGA TTAGATGCTT TGTATTAAT TGTCTTCA TGAATCATTT 600  
TATCTTTTAA TATGATGA AAAGTGTGCT CTATATTTAA AAAATTAAT ATTAATATAC 660  
40 AGACAGCAT ATACATGAA AGTCAGTGC AAGACCAAT CTTCAGAGT AAAAGCAATA 720  
TTCTTATACA TTTCATATTA AATTAGCTTC TATGTATTTT CTATGAGACC TGAAGAGGCA 780  
GATCCAGAT TCTTAAAGCC TTGTGTTGAC CATTGTCTTA GTTACTGCT GAAAAGTGA 840  
45 TATATTTTCC AGCATGCTT GAGAGCTGT ACTTTCGA TGTCAATTAAT CAGTTGTAAA 900  
ATATATGAGA TGTGTCTCT TGTGTACAT TGAAGAGAAA AAAATTTT TTCTTCACT 960  
50 GTAAAGAGG TGTGGCTGAC ATCAAGATTC TTCTGATAT TTATCTCAT GCTGTACAAA 1020  
GCTTAATGT TGTATGATTA TCTTACGTGT TGAAGAGTGC ACTGAGAAA CAAAATGTGC 1080  
AATATGCGTA AATTATATTT AGAATGTGT GAGAGCTAT TCTGTGAGAA AAGTTATAT 1140  
55 GTCTAATAG AGAAGTCTTA ATGAGCTGTC TGAATATGT AACTGAGTT ACAAGGTGAC 1200  
TTTAAATGAC ATACAGTAT TTGATGAAG GAGTGCAGC AATGTGAGCA TGTCTGAGAA 1260  
60 AGTTATCTTT CCCCCTCTT GCTGTGCTCA TTGTGTCTTG CAGAAAGAT GAGCTGATG 1320

5 CAGCAGACC GCGAGCTGTA ATTAATAATA ATTGAGCTA TGAGACTAGC AAGGCACTAG 1380  
AACTGGAAA GACCACAGAA AACAGAGAT CCAACCTTTT CATCTTACAG GTGACGAAAC 1440  
TGTGATGATG CACATGTATG TGTTTTGTAA CGTGTGAGCA CCGTACAAA ATGTAAATTT 1500  
GCCATTATTA GGAAGTCTTG GTGCGAGTGA AGAGAGACC AGGCGACTTG ACTCCAGATC 1560  
TGTGCGCTTG TCTACACAG ACACACAG AGCTGGGTCA GATTCCTCTC AGCTGCTTAA 1620  
CAAGTTCTCT CGACAGAAA GTGCTTACAA AGCTGCTTTC TGCGATACG AAGGTGCGAG 1680  
TTTTCTGAAC TCGACTGATT TTATTCAGT TGAATAAAAA AAAAGCTAT TCCAAAGATT 1740  
15 TCAAGCTGTT CTGAGAGATC TTCTGATGCG TTTCCTTCTT CAGAGGCAAT GTTTTACTTT 1800  
TATGCAATAT TCAATGTGTC CAGGANTAA AGTGAAGAA CAGCAGCTTT TANTATATAG 1860  
GTCTCTCTGG AAGAGAGCTA AATTGGAAG AGAAAGCTGT GACAAATTTT ATATTTCTAT 1920  
TCTTAAAAAA CACTAATCTT ACTTACAAA AGTCTCTTTG AGAATAGTT ACACAGATG 1980  
GCCACAGCAG TTGTGCTTTA ATAGTATAGT GCTTATACG ATGTAAATGG TTACTACTA 2040  
25 CTGCGCTTAA AAAAAAAG CAGCATATTT ATTGAAACA TGAGACAGA TTATATGTC 2100  
TTAACGATA TATTGTGCA CTTAATAAT ACAATTAAAA CTCTCTCTCT GCTCTAGTAC 2160  
CATGCTTAGT GCAATGATT ATTCTATCT ACAGCTGATG CTGTGTCTTA TTTTATAAA 2220  
30 TTTATCAGAG TGAATAAAAA AAAAAAAG AAAAAAAG AAAAAAAG AAAAAAAG 2280  
AAAAAAGAA AAAAAAAG AAAAAAAG AAAAAAAG AAAAAAAG AAAAAAAG 2320

## (2) INFORMATION FOR SEQ ID NO: 25:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 683 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

50 GGCACGAGCC TGTGTGTGTA TTTTCTCTGT GGTGCACTAC CTGACATGAG CAGCCAGCC 60  
TGAGTGGCTG AACAGCTTC CACAGGCTG CAGGTGTGTG TGTGTGTGAA AGAGAGAGGG 120  
GGGCGCAGAG CGCGCTTTTG AATGTTTGC CTGTCTGAACT TGTGAGACA CTTCGGAGTG 180  
ATTGTGTCT AATTTCCAC CTGCTCTGTT TTCTGTGACA TCTTGGAGG GAGCTATGTC 240  
55 CACACCATGC GGGGTGCTTA GAATATAAA AGTCCCGGGT CTGTCTCTCT CACTCTGCTT 300  
CTCATGGGGG AGGGAAAGAA TGGCTTTGGT GCTTTGTTTC ACACAGCTGA TGGTGTGCG 360  
60 GAAGTGTGTC ACAGTGAGCC TGTGTGAGG ACTGTCCACA GGTTCACAC TTGTCAACAT 420

5 CAGGCTTTTC TGTCTCTGAT AGGTTGAGC AAGATGAAA AGGAAGAAA AGAGCTTTT 480  
CTCAGAGCCA TTATATAAA TAGTAGGTGG ATTACATCTT CGTGCTCTTG GCCACTTTC 540  
CGTGTGCTC AGTGACATGT AGATGACTGA CTGCGATATC TTGTCAACAT TCCCTGGAG 600  
CAGCTACTTA GGGGAACAA GATGTAGTGC TATTCGCGAT AACAGATGAG ATTTTCCACA 660  
10 CTAATAAAAA AAAAAAAAAA AAA 683

## (2) INFORMATION FOR SEQ ID NO: 26:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2036 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

25 CTGAGAAAG AAGAGATTCG GATCTGCTGC AAAAAACAT ATATCATAA AGACTCATGT 60  
TATTGAAA ACAGATTGTC AACACATCA CATTCGATG AATCCTTAA AAGGAAGAG 120  
ACCTTAAGT ATCTGCAAT CTGAATTTCT ATTTATCTCT TCAGTGAATA TAGAAACAT 180  
GGTATCTGA TTATTAGAGA TATATATTTG GAATGTTTAC TTATTAATCT CTAATGCTG 240  
GTAACTGCA TAAAGTCTGT TATTAATAAC AACATAATTC TTTTATAAA GAAGAAAGC 300  
35 TTATTTTCA TTGACAGTGT ATAGATTTAT CTACTTAGTT GTGTTTGTCT ATTAGTGTTT 360  
TAAATTTTTT TTTAAGTGA GTGTTTGATA AATTTTAGA CCTGTGCCC ACCTGTGTTTT 420  
GAGTCTGTG TTGACTACAG GTATATAGCT CAATTTAAA ATCTTAAAG AAGAAATTT 480  
50 TATTTATAA AGAATCAATC NGTTGCAATG ATGAGGCTGT GAATCAGAT ATTATGTAAT 540  
AAGAGACA GTGCTTTTTT TTGTATTTAC CCAATGAGCC CCACCAATG CAACTGTGTTTT 600  
65 AATATAGAA ATATGATTAACA ATTTTAAAT CTCAGAGTAA AATCTATTTT ACTACATGCT 660  
TTTCCCGCTT TGTCTGATTT TAAGCAGTGT GTACTTGCA TCTCTACATT GTCTTAGGGA 720  
75 CAGTGTGTTT CTACAAATTT ATCATGTATG ATGTTTTTAT GTTCTTTTTT ATTCAATATG 780  
GCTTCTTACC AAGAACAGTA GGAAGAAACA CATGCACTGT GTACAGACA TGAACATATG 840  
90 CTGCTGATAT GTTGTTTTTT CACATGCTTT TGAGTTTTCA CTTTTTAAAC GAGAGCCAGC 900  
55 AAGCAATAA GATGTGCTG GGTGTGCTG TCCGAGGCGC TTTTTCAGC GAGCTCTCAA 960  
ATCTGTGTA TTGAGGTTTC CTTTGTGTA CTCAGATTG GAGCTACAGC TGGGCCCCCC 1020  
TCTCTCCAT TCGTTTGAG AGACACTGAG GGAACAGG GTTCTTTTG AGGTGCTCTT 1080

255

GGCTGCTTTT TACGGATGCG GAGCTTCTT GTCCTCTGCG AACCTCTGTA 1140  
GGTACGCGCG GTGCGAGCTT GTTGGATGTA TTTCCCGACG GCGCGCGTGA 1200  
CTGGCGTGA TCCAAAGCA TCGACACGA AGCGCGCGGA GCGAGAGAA AGCGCGCTG 1260  
GCGTCTTGG TCCCTCGAG ATGTCTGTAG CATCTAGCTC GAGCTTGGCG CTGGGGAAGC 1320  
AACCTGACGA AACCGCTGAG GTGTGCGCTT TACAGAGGA ACCTCGAGGA GATTAATTTG 1380  
GACATGAGC TGTATGACA CTATGTGGA TTTTTTTT TTTTACAGT CATCGAGAT 1440  
GTTTGAAG TGAATTTT TTTTGTGTA TTCTTTTC TTTACTTAA GGTGAATGTG 1500  
TATTTCTG GAGAGATTG GAGAGAA GAATGTGTA TATGTGAA CAGAAACT 1560  
CGTCCCTTAT TAAATGATA TCTTCATGTA TTTATGCGAT AATGTAGCT GACTTTTAA 1620  
AAGCTTCTT TTTGTGATG CGCTGTGAG GCACTGTAT TGTACATGA TCCCTTCTG 1680  
CGTGTTCG TGTATTAAGT TAGTGACAA AGAAATATT TTGCGCTAG TCAATGTCC 1740  
AACATGCA TATTTTAA ATTGTGTA TAGTGAGGA GCAATTTGT TACATGTA 1800  
AGCTTACT ATATACAT ATACATAT TTAGAGATG TGTCTTGC AAGGTGACG 1860  
TTTTCGAG TTTTACAG TGAAGACG TTTGTGTTA AACTGTAC AATGTATTA 1920  
TTATCATTT CCTCCATGT AACTAGAT CATGTGTA TTTGATGCA ACGTAAAT 1980  
GAAATGAG GAAATGAT ATATACAT TTTGTACAA AAAAAAAA AAAA 2036

35 (2) INFORMATION FOR SEQ ID NO: 27:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 717 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GGCAGAGAT AACATAGCA CATTATAT GTATGTAC TTTAGCAT ATTAAGAT 60  
AACATGAA TGAATTTCC ATTGAGAG AAAATAGTGG GTTTCATGCG CTATTATTTG 120  
ATTCTGAG AGAGAGAT GCGACACA TTCAACCA GCGCTGATAT GAGATTAATC 180  
CTCAGAGCA GAGCCAGCA CAAAGAGCA ATGTAGAAA GTTACATTTG GAAATTTCC 240  
TGCAGCTTC GGAATGACA CTGCAAGCT GATGCCGAA ACTGCCAGG TAAATTTCT 300  
CATTAAGCT CTACCAAGC ACTTCACT CCCCAATTA ACTAGTGG TTAGATATC 360  
CTCTTACCT TTATCATTTA GTTAGAGAT TGCAGAAAA CTTCATCTT TCCATTTAA 420  
GCGCTGAT TCTCTGAT CTGTATGAG AACCATGAC ATTAATCTGA AACATGAT 480

256

AATAGAGAT TCTTATCTT CTCCGACG CTCTGATAT AAAAAATC AATTTTCA 540  
GCTGATCTA TACATCTTA AATTGAGAA CATTAGATA AATTAATGC ATTAGAGGA 600  
GATTAATAG GATGCTGC GTATCTCTT AAGTTTCC TTTTACAG AATGTGAG 660  
GTCAATGAT CAGCATTAG ATTAATTAAT GTTAACTTA AAAAAATTA AAAA 717

10 (2) INFORMATION FOR SEQ ID NO: 28:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 495 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTCGCA CGAGAGAT CTATATTTA GTTGAAT GATTTGAA GATCTTCTC 60  
TATCTTTT TCTCCACA TTATCTTA TTTCTCTGT CTGTGAT TGTGATAG 120  
AATGACTG TGTTTGAT AATTAATGA GAGTGAAT TTTTCAAT CTAGAGAA 180  
AGCGATTTG CTCTACAG CTGAAGCA CCGCTGGTG GCTGGGCCC TGTGGAGT 240  
TTCTGGGCA TTGACCTTA CAGATGAG TGGCTTCA GAAAGCTG CAATTAATA 300  
TTATTTTAA AAAAGCTCC TCGAGAAAT GCAATGAG GATTAATG CAGTATTTG 360  
AACCTTGA GATTAATAT AACCTGAG AGAGAACTG AATCATTTA GAGCTTGA 420  
ATGATCAT GTTCCACG TAGAGATCC TCCGACAG TACAGCAT CAACCATTA 480  
CAGAGAAAT CAGG 495

40 (2) INFORMATION FOR SEQ ID NO: 29:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 556 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAGTAACT CATATTTAT TACGGAATG CAGCGAAA CATTATGAG AATGCCCTA 60  
GACACTCT AGAGAGCTG CTAGAAAGC AGACAGAC AACGCTTAA ATGAGATGG 120  
GCACTGAG CTCTCTGT GCTTATGCT AAGGCTGAG GAGATGTAT CATCTGCA 180  
CATTAAGT GATATCAT GAGCGACA CAGCATTTT CTGGTTACT TACAGAGAG 240



ANTAGAAAGC AGGCAGATCT TTACAGACGC TCTTACTCTN TTCCAAAACA ATCGAAATGC 300

5

CCACATGTCC ACUACACAGT KTGTTGTCTG CTTGTGCCAT GAGACACAGT GTGCTGAGGC 360

10

GTACAGATCT CCCACACTCA AAGGAGCCAG CAGATACAGG GCTGCACACT GTGTGATTC 420

15

ACACATGTGA CATTCTGAC AGGCAGATGC TCGATGGCAA AACGAGCATC GGGCTGAGAG 480

20

GACTGCTGAG AAGGCGACGC GGGCTGCTGG GATGTGGGTT GATTGTAGCA GTAGCTCATG 540

25

GAGATGTGAC CTCAAA 556

(2) INFORMATION FOR SEQ ID NO: 30:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 434 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTAAATGGTG ACTGTGGCTT TGTGAGACA GGGCCCAAT GTTAGTGTG AACACACAT 60

35

GCACAGATG AGGAGCATG CAGAGTGCTG AATACTGTC CTGCACAGAT GTGTACATG 120

40

ACTTTCTTTT CAGCTATTT CTGTGGCTG CCTTTGAGA TAGAGCTTTG TTGATATTTA 180

45

CATTAAACCA AATTGTATTA TTATGTCCA TTCTGACATG TTATTAGCA AARGAAMAA 240

50

GAGTAATCTT ACATCAGCAT CTTTATGCA TGCTAAAGCA TTAAATAT CTTTTGGGGA 300

55

ACATGTTTTG TATACATAAA TGTTTAGATA GAATATTTA TAGAATNCTC TATGTAGATA 360

60

TTTATCTCCC TATGTATATT TTAATCTAGA TGTGTCTATC TTTGTATGCA TATGAATGC 420

65

TATGAATAGT GAGA 434

(2) INFORMATION FOR SEQ ID NO: 31:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 715 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

70

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCAGCGGTCC GATCTCAGAG CTCGACACT ATTCCGAGCC ATACACAGCC TGTGTGTAGG 60

75

AAGCTACTC CCAACTAAG CCCAGATGC AAGTTTGGT TCAATGGGGG TTAGACAGCT 120

80

ATGACTATCT CCAAAATGCA CCTCTGGAT TTTTCCGAG ACTTGATGTT ATTGGTTTG 180

CTGGCTTAT TGGACTCTT TTGGCTAGAG GTTCAAAAT AAAGAGCTTA GTGTATCCGC 240

5

CTGGTTTCAT GGGATTAAGT GCTTCCCTCT ATTATCCACA ACAAGCCATC GTGTTTGGCC 300

10

AGTCACTGG GGAAGATTA TATGACTGGG GTTTACGAGG ATATATATGTC ATAGAAGATT 360

15

TGTGGAGGA GACTTTCAA AAGCCAGGAA ATGTGAGAA TTACCTCGA ACTAGTAGA 420

20

AACTCCATG CTCTGCCATC TTATCAAGTT ATAGETAAC ATTGGAACTC CATAGATAA 480

25

ATCAGTATTT CTACAGAAA ATGGCATAGA AGTCAGTATT GAATGTATTA AATTGGCTTT 540

30

CTTCTCAGG AAAACTAGA CGAGCTCTCT GTTATCTTCT GTGAATATCAT CCTACAGCA 600

35

AACTAAGCTG GATTCCTTTC ACTTAGAGAT AATGTACAGG CCTTAGAAGT CCTATTTCTC 660

40

ATGTTGCTAT TTATGTACTT AATTAAGACC CAGTTAATA AAAAAAAAAA AAAA 715

(2) INFORMATION FOR SEQ ID NO: 32:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 486 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GAGCCATGTC GGGGAAAGG GAGCTTCTCT CTACTTCTCTG CCACAGAGCC TGTCCCAACA 60

120

CAGTTCTGTC CCTCTCTCTG CTGGAGGACC ACTTCTCTCC CGAGTCTGTG ATTCCAGCCC 120

180

CAGCTCAGCC TCAGACTG GCGCTCTCTCT CTCTCTCTCT GCGCTCTCTCT GCTTCTCTGA 180

240

GCGTCTCTCT TCTTCTCTCT TTGAAAGCA ATGCAAGCTT CCGGGATCT TCTGCCACT 240

300

CGAGTACCA TGCGCTTCTG TCTTCTCTCT TACGCTCTCT AAGGGAAATTG TCTATCTCTG 300

360

GTGTCTCTCT TCGTCTCTCT ACCTCTCTCA CCTGTCTCA AGCTGGCATC TCGCCCTTCA 360

420

CTGCACAGAA GCGATCTCTC ACCACTCTCC TTATACAGGA GGAAGCAGCA ACNTGAGAGA 420

480

ANCAACTAT AGGGCTTACA ANCATGTCTCA GCTCTGATCC CGAAGGCAAA AAGNATCTTT 480

486

GGGAC 486

(2) INFORMATION FOR SEQ ID NO: 33:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 725 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

5 GTTCCCTGCG TAAATATAG GTTATGCCA GAAACAGAT GTCAATCTT AAAATTAAC  
TTTCCCTTT AAACCTTTT AAGAGAGAG ACCACCTTA AGATGCCCC TAGGGTGAAT  
ATGCTCTAA TTCAATTTAT AAAATTTAT CTGTCTTCA TTTAACCTT TTGGCTAATAT  
10 AGTCAGAAAT GTCCATAATA ACAACCTAT TTGATTTAA TTAGGGAG AGTAAGGGA  
AGAAAATGA AAACCTAGTC TTATATAGG CTCACAGAT ATTAGGCTT AAAGGCTTT  
TCATGTTTA TGAAATTTG TACTACTAT TTTTAAATAT TCCGTGTTT GATGAACGA  
15 TCTCTGGGA AATTGTAGG TTACATAGC ATTTCACCT GATCCCTTC AACCTGAGAT  
CAATCTATA ACCAATGAC AACCTCTTC TTGGGTTAC TGTCTGTGA AATGACAGT  
20 CAAGTTCCC AGAAGTCGT TGTATGAT GATGACAGT GCTTTCTTC GTGGAGAG  
TTCTGCCCC TCTTAATTT GGTGTATGT CTCACAGTA TCTAACTTC CATCTGATC  
TGATGACTT ATCCATAGT TTAATGTAT TATGATTAAT GTTGAATATC TTGCTTGAAG  
25 GTTCAACTT TTTCAATGA TGAATAAAA GTTTTCTCT GCTATMAAA AAAAAAAAA  
AAAAA 725

(2) INFORMATION FOR SEQ ID NO: 34:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

5 CACACAGAT GCTGCCCTCA GACCTGTCA TCTGTACCA CATTGAAAG CTCCTGCTCC  
45 TGCAGATAC TGAAGATAG AAGCATGCTC TGAATATTT CCGAATAC TTCCAGATGC  
CTCTGCTTT TATTGACTCT TGTGAATC AAATGGAAG ATCTCAGTC CCACTGAC  
CGAAGTGA AAATATTC AGTTCATCC CCAAGAAC AACACGAT ACATGACTC  
50 AGAATCTTA TAACTAGAT GATCTTTTC CATCTGTACA TTGTGTGCA CATGCACTC  
ATCACTGAC GTGCCAAT CCTGCAGAG CAACACCTG TGAATATTC AGGTATTTT  
55 CTACATGTC AGCTTAGAT TACCTGATA TCACTTACA TTCTACIAN AAAAAAAAA  
AAAAAAAA AACTGTA 437

(2) INFORMATION FOR SEQ ID NO: 35:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 943 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 GGCAGAGCT GAAACAGGA CTAAATCCA GAAATGAC ATTGTATAC ACCTAATAC  
10 AGAATAGCT ACCTGTGAA GATTAAATAT AAAATGTTG ACGATATGA TGTATGAT  
GAATGACAG ATGCTATAG TGTACATTT CATTATCTT GTTAGATTA CTTCTCTTA  
15 TCTTGTCCA TTACATGCT ATTTCTCTT CAGTGTAT TTCTAGTGA CAGATCTTA  
ACATTTCTTA CACCTGCGA GAAAGGAGG AAATGTGTT TGGGCTGGCT AACTAATTT  
20 TGAATGAAA TATCATAGA TGAATATGA AAAATGAGA CAGAAATGT TATTAAGAAA  
AAAATATGT TTTTATACC ATTGACTG CTCCTTAAT AGCTACAG AGATTCAGT  
25 TTAACATAC TTATAGTGA AAAATGTG CCAATAGT ATGTCTTCA AAAGCGAAA  
TGTGCTTGA TCCCTTAAAG CTAAATTTG GTCAATTCG ATCGAGATG TTGTAGTAT  
30 TGCATTAAT ACGACCTAT TTTCATAGG TGTATTTT TGTATGACT TTTTATACC  
ACTATATGT TGAATCTTT TTGTCTATN AGGTGAAN ATGACATGC TAAATGAAA  
CAATTTACC AATGCGAAA CTAAAGAGG CAATTAAT ATGAAGAAA TTATGAATT  
35 ATGATATTC AACCTTGA TTAACTTGA GATTAATAG TAAATATG TGGAGAGAT  
TTGTGTTTC TGAATTTT GTGATTAAG GAATGATGT TACCTTTAT TTGTAGAAA  
40 AGTGAATGT ATGTGTTTA TTAATATTA AACTGATCC ACGAATAAAA AAAAAAAAA  
NAAAAAAAA AAAAAAAAA AAAAAAAAA AAA 943

(2) INFORMATION FOR SEQ ID NO: 36:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 604 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

55 GGCAGAGA AATTCATGC TGTATGACT CCAACAGAG GATGAGCTTT CCACTGAT  
60 GAATCTTAG TGTGAGTGC AGTGTGTTG TTTCATAT TCTTCTAAT TTTTCTTCA 120

TTGATCTTGG GGAATTTTCT TTGTTTCTCT CTGTGTTTGC CCAGTTTAA TAAACACAGG 180  
CCCAACAAA AACATAGCA TTCTGACAA TAGGGGGCC ACATTGGACC CAGTATGTCA 240  
5 CTTTAATGCA CTTTACAGAA AATCTGAAT GCGAAAAATG ACATAGGAA TGTATCTCC 300  
AACAATTTTA TCCCATATAA TGSTGTGTTT TCTTAATTTT GTTCTTCTG CCGAAATGTG 360  
GCTTTCMAAT TAAATGACC TTITCTCTCT TGAACITTTT TTTTGTGACT TGTATATTA 420  
10 AGGTTTGGCA AAGATTTCATA ATTCTGAGAG AGGTTTGGCA CCAGAGATA CAAGAGATC 480  
TGAGTAGTAA TCTTTTTCAT GTGCTTTTAC AGCCAGCTAC ATTAAAGGAT GTATTAGTTA 540  
15 CAGAAATTTAT ATGTCTGTCT ATGTCTCTCT ACTCAATATA GTACATGCTT CCGAAMAAA 600  
AAAA 604

20

## (2) INFORMATION FOR SEQ ID NO: 37:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 349 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGATGTCCC GGGAGCCGCC AGGCGCTGCC CCTAGAGGG ATATCTTTTA CCGCTGCTTT 60  
GTCCACACC TAAACCCCCA GCTCTCTAGG CAGTGGGCAC ATGGCAAGGG CTCACATGGG 120  
35 GGCACATAGA GCAATTTGGGG GACTGGGAGT GCTCACTTTT GACTTCTGTC AGGTGCGGGG 180  
AAAACCAAGT CACTGATACC AAGGTTTACA TATTTCTTGT CTTCACTGTTG CTGATCTCTC 240  
40 CCTGCTCTGG TCTCACCAGG TATATGCCAC CACTTCTCTG TCTAAATTTCA GAATAGAGT 300  
CACATCAGGA GAGCACTGTC CCGAGANNA TCGAAACGGG TTGGAGCA 349

45

## (2) INFORMATION FOR SEQ ID NO: 38:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 672 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GTAGTGGTTG CCGTTGGCGG GATGGCGAAG ATCTGCGGCT TTGAAAGTGT ANAAGGACCC 60  
TGGTAGCCGG TCGTTGTTGG TTGTGTGATTT GTATCTGTTT CTACAGAGCT GATGCTGACA 120

60

GGAAGGCG CAGCGTCCAC AGGCAAGTAA ATAGTAATCC CCGACCAAGT TTCTCTCGGC 180  
TTTATCATGT CACCACATGT GGTATATGCG TTGTGTTCTG CCACCTTTGC CGTGAAACAT 240  
5 TTCCAGCATA ATCAGATGCG GCGTGGCGCA ATATTCACGA TTAAGCCTGG CAGTGTGGG 300  
GCTGATGTTT CAGTGGCTGC GSCACCGTTT YTGCGGTATG TTGCACACCA GGTCTTTTAA 360  
ACAGTTTTCG SAGCGGTTT AGGTCAMGG GTTCATATCC GTCGCTAGCC TGTCTCTTAG 420  
10 GTTCACGCGG AGCATAGCA TTAACATCT CATCAATTTG CTTCTGGCTG GCGCTATCMA 480  
TACTTTTCAG CATATGTTTA CCGTGGCGCA AACGGTTAG CTTTTCGCC ARCMGATCAT 540  
AGCAATGCG CTTAATGAGA TAATCAATA CACACACAG TACGGCTTCA GACACCGTTT 600  
15 CCAATGCTT GCGTGGAGTG GTAAACACCA CCGCGCGGG ATATGCGCC TGCACCAATT 660  
CATGCAGTAA AT 672

20

## (2) INFORMATION FOR SEQ ID NO: 39:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1908 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGAGTTGATA TTTTATGAAA CAGTAATTTT ACTTTTAGG AAATGGCTA GCTCTTTGAC 60  
35 TTAAGAGCTG TAGAAGCTC AACATTTCTT TGTAGAGAC GTTGTCTTTT TTGATTTGTA 120  
CAGGTATAAA AACATGCTT TTGTGATTT GTATAGGTTT AAAAAAGAA TAACTGTATG 180  
40 CAGGTTTCAA AAGCAATGT GCTTTAGCA TCACTCATAA GATGCCATG TACTTTAGG 240  
CATTTTATTT TCGTTTAA AAGCAATCA GCTCTTCTCT TCTGACTGT AACACATAGC 300  
360 CCGAAGGAT GAGATTTT TTCAATGGT TTTTATGTT GTTATGTTT GGTTTGTAC  
45 CCGAGCGCAG TGTGTTCCG GACACTGAC TCTGCTCTCT AATGAGACA AAGTTAGAAA 420  
TCTGCCATA ACCTAATAA ATTTAGAAAT GAATTAATAA TGTGAATG GGTAAAGTG 480  
50 ATGATGATAA ATAGCATGC AAGAAACAG CTCCTTCAT CAGACTGCG TACTGTTTC 540  
TTCTGTAGG ATTTGTTTG GAGAGCTTC TTGTTTCTT CTTTGGGG TATGTTCTCG 600  
55 TTCTTTATA TGTTTGTAAC ATTATGAGA TATATTCAC ATACCTTACA ATTCATTTAT 660  
TTTAAGGATA CAATTTAGTG GTTTTTAGTG TATTCAGAA GTTGTGTAC CGTGACACA 720  
GTCAATTTTA GACATTTG TTAACCCAAA AAGAAACCT GTACCTTGA CGAGTCAGCT 780  
60 CTCATTTTCT CCGATGTC CCGCATCC AGCCATCC CAGCGCCG GAACACTAA TCTATTTCTC 840



## (2) INFORMATION FOR SEQ ID NO: 42:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1983 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

5 GGCACGAGAG GGGCCGAGCC GACAGATGT TCTTCTCTCC TCTTCTCTCC GCGGGGGGAG 60  
15 TAGTCTCTCG AGCTCTCTCC GTACAGCTTT TCGGGAGCCG GAGTCTCTCC ACCCCAGACA 120  
TCGACGAGGG CCTCTCTTTA GGATCTATT CCAAGGAAA AGAAGATGAT GTCCACAGT 180  
TCACAGTCC AGGAGAGAT TTTGATAAT TGTTAAGCTGG AAAGCTGAGA GAGACTTTGA 240  
20 ACHATCTCG ACCACTCTG AGGCGAGGGA AGACTCGAAC CTTTATGCT CTGCTCTAGG 300  
ACTTCCCGAG CGTGTCTCTA GTTGGCTCTG GCAAAAAGGC AGCTGGATC GAGCGACAGG 360  
25 AAAACTGGCA TGAAGCGAAA GAACATCA GAGCTCTGT TCGAGCGGG TCGAGCGAGA 420  
TTCAAGACCT GGAGCTCTCG TCTGTGARG TGGATCCCTG TGGAGACCT CAGCTCTCTG 480  
CGGAGGAGC GTTGTCTGTG CTCTATGAT AGCATGACCT AACCAAAA AAGAGATGG 540  
30 CTGTCTCTCG AAAGCTCTAT GGAATCTGGG ATCAGAGGGC CTGGCAGAAA GGAGCTCTGT 600  
TTGCTCTCTG GCGAGACTTG GCACGCCAT TGATGGAGAC GCGAGCCAT GAGATGAGC 660  
35 CAGCAGATT TCGCCGANTT ATTGAGAGA ATCTCAAAAG TCTGTATGAT AAACCGGAG 720  
TCCATATCAG ACCCAAGTCT TGGATTGAGG AACAGGCAAT GGGATCATTC CTCAGTCTGG 780  
CCAAAGGATC TGAAGAGCCC CAGTCTCTCT TGGAAATTTCA CTACAAAGGC AGCCCCCATG 840  
40 CAAAGCAACC ACCCTCTCTG TTTGTTGGGA AAGGANTTAC CTTTGAAGT GGTGCTATCT 900  
CCATCAAGGC TTCTGCAANT ATGCACTCTA TGAAGGCTGA CAGTGGAGGA GCTCCAACTA 960  
45 TATCTCTAGC CATCTGTCTCT GCTGCAAGC TTAATTTGCC CATTAATAT ATAGGCTCTGG 1020  
CCCCCTCTTG TGAATATATG CCGACCGGCA AGCCCAACJA GCGCGGGGAT GTTCTTGAAG 1080  
50 CCAAAAACCG GAAGACCATC CAGTTTGATA ACACGTATGC TGAAGGAGG CTCATCTGG 1140  
CTGATGCGCT CTGTTTACGA CACAGCTTTA ACCCGAAGAT CATCTCAAT GCGCCCACT 1200  
TAAAGAGTGC CATGATGATA GCTTTGGGAT CAGGTGCCAC CAGGTGCTTT ACGATTTCT 1260  
55 CCTGCTCTTG GAACAACTC TTGAGGCCA GCATTTGAAAC AGGGAGCGT GTCTGGAGGA 1320  
TGCCTCTCTT CGAAGTATAT ACAGACAGG TTGTAGATATG CCAAGTCTCT CATTTTAA 1380  
60 ACHTTGGAAA ATACAGATCT GAGAGAGAT GTACAGCTGC AGCATCTCTG AAGAGATGG 1440

TAACTCATCC TAGTGGGCA CHTTAGACA TAGAGGGCT GATGACGAC AAGATGAGAG 1500  
TTCCCTATCT ACCGAAGGC ATGACTGGGA GGGCCAGAG GACTCTCAAT GAGTCTTAC 1560  
5 TTGCTTTGAG TCAAGACAT CCTTAGTCA GATACTCAA AATGCTCTCA CTCTGTCTTA 1620  
AATGGACAG TTCAACTTAA AAGTTTTTGG AATAAATGGA TCAAAATCTT TTAACCGAGA 1680  
CAAGGATGG TATTTAAAA TGTAGACAC AATGAATTT GTATGCTTG ATTTTTTTTT 1740  
10 CATTTACAC AAGATTTAT AAAGTAAAG TTAATATCTT ACTTGATAG GATTTTTAAG 1800  
ATACTCTATA AATGATTAA ATTTTAAAG CTCTCTATC ACTTTTACA GTATATGTTT 1860  
15 TTCAATGGA AGCAAAATG TAACTCAGT TTGTGATCT AGGACATGA GCAAACTGAA 1920  
AATTAATG CACTGTCTAG AAACATAAA TCCAAGTGT TGTCCAAAA AAAAAAAAAA 1980  
AAA 1983

## (2) INFORMATION FOR SEQ ID NO: 43:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1406 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGATGATCA CTTTGAAGAC GATTTTATC CTCTCTCTCC AGCTAAGGC CTGAGGTGA 60  
120 ATAGTTGAAA AAGACTCTAT AGATATGAC ATTTCTTCAA GAGAGAGAGA AGTCAAGTCT  
TTAAGGCTTA ATGCTTAAGC NCTTGCTCTT AACTGACCT GCGATAACTA CTTTAAAGAA 180  
240 ATAAAAAATT CCAATCAAT ATTCTCAAC TGAAGTTTA GTGGCAGCAG TTCTATTTGC  
CTTCACTTA TCAAGTACT ATTGTAGAAA GTGTAGACGA TACTGACTCA ATTTCTAAGT 300  
360 CTGATTTTGG CAATTTTITA TGTACTTTT TAAATAGCT TCTTACGTGC AATCTGAT  
TAGAGTAAA GCGCTGTGT AAATAAAGG CTCAGGCAAA ATTGTACAGT GATAGCAACT 420  
480 TTCCACACAG GAGGTGAAA ACAGTAATGT GGTACACAG TTTTITTAAC TGTAGAGCA  
540 TCAGTGGCT CTTTATATTA TGACTTAAACA ATATTTTAAA ACAATCATTA GTACGACAT  
600 ATTAAGGCTT TCTAGTATG TAAATACAC AGCAATGATC TTGTGCTTTT TCAATTTATT  
660 GCTAGATGTT TCCCTCTGG AGTTTGTCA GTTTCACACT GTTTCCTGGC CCAAGGTGAC  
720 TGTTTCTGGC CTTTGTATAT ATGCCAAACC ATTGCTGGG AGTCAGATG GTTCTTTAAA  
780 AAAAAAAAAA AAAGACAT ACCTGACAGC TCACTTTTC AATTCTTATA TGTACCGAGG  
840 GTACAGTGT GTGGGATGAG GTTCGATACA GAGTATTTTA TTGCTTGTCA TGTAAATTTA

AAACCTGTGA TTAACTCTTT TGCATCCCTT TTAGATATAA TTGCTCTTGG CAGAGATGAT 900  
TGTGCTTAT TTTTTCAAA ATTGCTGTG AACAGCTGA TGCACACAG CAGCTTTAT 960  
CTAATGACTT AACGCTATCT TAAATGCTT CTTCAGAGTT TGTGTGTGAC TTGTAAAAAG 1020  
CTACAGAGA TATTAATAAA ATCTATTCAC TTAACTTAT AATGATTTAT GAAATATAAA 1080  
CATGATGCAC AACTTTTGT CTGTGTGTAC CTATTAATAA AGTTTGTCTT TGCATTTCAA 1140  
TGTAAAGAC TGAAGACAT GTTATGTGT TAAATTTGT TGTGTGTGA GAAATTTTGG 1200  
TGTATAGAAA TTAAAGAGAC TTACAGAGAA GATTTTTAG TTAGAAATAT TGCATGCCAA 1260  
TAAATATAGA AATTAATAT ATTAAGTTT TAACTGTGAT CAGTGGAGAT CTGTGCTTAG 1320  
TATGTATGTG CTATATATGT AACACAGAG ATTTTTTTGG CTATTTAGCG TACCAAAAG 1380  
GACCGATTA ATGGTTTGA GACTTG 1406

25 (2) INFORMATION FOR SEQ ID NO: 44:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1391 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGCTGTAGG GCGAGACGCC AGTCCGACAC AGTGTCTGCT CTTGTGTGGG GGGCTGCGGC 60  
CCCGGGGCTTC GCGAATGACA GTGAGCTGGA CATTCTGTGG GGAACACAGA CCGTTATGGA 120  
CGAGAGCTGG TATGCGCTCT GCGTGAATGG TTAACTGTGGT AACGACGGGG TGGCGCTTGG 180  
GGTGTGCTGG GGAATCTTGG AACAGACTGG CCGCAGCGGA GCGGTGTCTG AACAGCAGAC 240  
CATGAGCAAT TACGCGACT TCCAGATGCT CAGAGCGCTG CTGCATGCC GCGCCAGACT 300  
ACTGTACAGG CTGATCTTCC AGATTTGGCC CTCCCGGCGAG GCACTACTGA TCGAGAGTTA 360  
CTATGCTCTT GATGAGGCTT TTGTGTGGGA GTTGTGTGGG AAGAGACTGT CCAAGCGCAC 420  
CAGAGAGAC CTGAGTAGCA TCGACAGCAA AACAGGATTC AACTTCAGAA GTGTGCGGAG 480  
ACATTTAGAC AACTTTAAAC GGGTCTTCAA GGTGTGTAGG GAAATCGGG GGTCTCTGTGT 540  
GAGCAATAT TACAGACAT TGTCTCTTTC TGAACGTTGG GCGAGGACT ATTCAGCAAT 600  
CGTCTCTTCT GCTAACACAC GCTTTGAGAC AAGGAGAGAA AAACGTCAAT ATTCAGACTT 660  
CGGTGACTTT GCGTCTTGGG CTGAGCTGAT GATCTAAAC TGAACCTTGG GACCGCTTGA 720  
CTCAGAGATG GATGACATGG AACATGACTT AAGACAGAA TTTCTCAGGA CTTCAGAGAG 780

CTCAAGGTGC TATGTGCTGA CAGAGACTT CTGAGCTTGC AACAGGACT GGTGTGACT 840  
GCTTCCCGGG AAAGCTGGGC GTCTTCTTGG AATGTGAGAC CAGTTTGAAG AACGTGTCCC 900  
GGGGGCTGTGT GAAATGTGCG CAGAGCTGAC CCAATATATA GATGTGTAGG AACGTGTGT 960  
GAGCTGTGAG GAGAGTTTGG TGGAGCTTG CCGCTTCCAG CAGTGTCCAG TGCAGCAACT 1020  
GGGTGTCTTC CTGAATCAT ATTCAGCTTC TTGTCAATCC CTGCATGCT TCCACACAA 1080  
GCGCTCTTGG GAGCGCTGCA TGGGCAACCT CCGCGGCTGCG CTGTGTCCC TGTATCTAGA 1140  
CTGAGTGTGC TCCCAAGTTC GCGCCAGACT GACATTAAG TTGCTGTAGG TTGTGAGACT 1200  
GCTCTGTGCT CCGGGAGACA AGTGGGGGGC GTTCAGATGT GCTGTGTCT GTCTTGTAGC 1260  
AAGTGTGTGC CTGTATAGAG GATGTATGTG TACATGTGCT CTTTGTGAGC TGAAGACTAT 1320  
CTCAGGGAA TGTGTGTGAG CTGCAGCAAT CAGCCAGAG AGTGTCTGA CAAAAAATA 1380  
AAAAAAAAA A 1391

25 (2) INFORMATION FOR SEQ ID NO: 45:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1569 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GGACAGATGG GAGATGTGTC CAGCGTGTGC GGGGATGTGTC CAGAGGATTC TTCTGCCCA 60  
GGGGGGGGGG CTGCGCTAGCC TGCAGACTGT CCGCTATGGC TGCAGAGCTG TTACCGGCGCA 120  
CGTGTGTGTC ATGCAGTTTC AACCGCAGAA GCTGAATGCT GTTACTGAA TATATCCGCC 180  
GAAACAGGCC ATTCAGCAAT CATGCTGTGC ATCTCTGCC AACCCGCCAC AAGAGAGAA T 240  
AAGCTGTATC AAGCTTGTTC GCGGGAGAA T AAGACAGAT TTTCAGACA ACCGATATAT 300  
AACCGTGTGC CAGATGTGTC CTGTGATGTG AAGAGACAG CTTCATTATG CAGACAGAC 360  
TCCGGAAGCA CAGATCTTGG ATGAGGTCT TCCCAACCA GTTCTTGAAG GCGCTTCTG 420  
GAGATTTCAA AATTAACAAA TCTGTGTGCC CTTTTTTGGG GCGACAGAT GTGTGCTGTG 480  
AATGAGAGAC CCAAGTTCAA GGAATGTGTA CCAATTTTAA GGAATGTGTC CATTCTGTGC 540  
GCTGTAGAT GAGTGTATGG ATGCAGCAAT CTTCAGACAG CAGGCTTTT TCAACTATTC 600  
CAGAGTCCCC AAGCTGTCCC TGTGTGAGGG GAGATGTGTA GAGAGCTTGA CTTGCTGTAC 660  
AAGCCAGACC CAGTCTGTGC TCCAGACACA GCGGCTGTAG CTGACAGACC TTGTGACACA 720  
GTACATCTGA GAGCAGATGG AAGAGAAAT CTGTATATTC GGCATATGG AACCCAGATC 780

CTGACACTGT TCGGACCTCG TAGCAGCCT GTTATGCCAG CCTTCGCCAT AATACACTC 840  
TGGTTATTTG GCTGTGCTGT COTCAATGGG ACATGTGGAA GAACTTGGGG TCGGGGAGTG 900  
TGTATTTCAC TTGGTTTATCA CTAGTAATCA TATTTGTGAGG TATATGGGCA GTTGGAGATG 960  
CAGAGGATTC CATTTTCAGAT GTACATCACG GCTTGTGTCG TTAGTTTTCG CAACTTGGGA 1020  
CCTGATAGCA GCAAAAGTCT TCATTTCTCC AGGTCCCAAG CAGAGATCCT GAAAGATGAG 1080  
GCTATTTATG CCTGCTCTCC TTGTGACTG CTTCTTTGCTG CAGGGCTCC TGAGCCCAAC 1140  
CCTTTGGGGC ACACCTTGGC ACTGCCAGAG TAGCTCAACC AAGCAATTGT GCTGAGATG 1200  
GCACCTGGTG AAGCCTGCTT GTGTGCCAGG CTTTGTGCTG AGTGTGTGTA CATGTATTAG 1260  
TTCTTTTACT GCTGACCACA TTGTACCAT TTCAACAGAA AGGAGCAGAG AATTAAAGTG 1320  
GCTTGTCTAA GGTCAATGAG TTAGTAAGTG GCAGAACAGG GACTTGAACC AAGCCCTCTG 1380  
CTCTGAAGAC CCGTCTCTGA ATTCTTTCAG TAGAGCTTCC TCATCAGGTT ACCGAGAGT 1440  
GGGTCCCATC CACCATCCAG GTGTCTTGG ATGTATGTTT TCACCCCTCG AGGTGTACGC 1500  
TGTGAANAAT TTGGAGACAG TCGTTTATTA TAAATATGAA TATATTCTAA AAAAAAAAA 1560  
AAAAAAAAA 1569

(2) INFORMATION FOR SEQ ID NO: 46:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1924 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GGGGCCCCCG WGGKATTTTT TTTTTTTTTT TTTAATTAGG ATATGCTTT TATTAACAG 60  
AATGAAGGT TCATTCCTCC TTCACTCCT TCTGTGTGT TTCTGGACA CAGCTCACT 120  
GATCTGCTA GAAAGGTGT CAGTCTGCTT GTGGCTTCCC TCTTGTATG ACTGACGCTG 180  
TGTGATGCT TGAGAGATAT CTATCACTT CATGTGATG AGCACTCCAA TATGACCCA 240  
CATCATCAT TCTTACTTAA AGATATATTA GAAAAAGTAA ATATAAAGA CAGGGTATA 300  
AATAAAGGT TTGAANAATC TAGTCACTT CAAATTTTAA AGATTAATAA TCCAGAGATA 360  
AAGATTGGGG GTAGTTTACA GCTAATAAAA ATAGAGAGAA ACTTCATGTT GGGGGGGAAA 420  
TCTAANAATTA TTCTTACATA AATATATGAG ACACCTGAT TAGATGAAA ACTGTATTTT 480  
CTTTAANAATG TAAAGAGCTG ACTCTCAGTT TCACCACTCT GAGCAGAGT TTGACTGCCA 540

CCCAAAATAT ACTATCCCTT ATGTGAAGGT ATGTGACAC GTTGACCTCA CCAATGAGT 600  
TTTACATCA GCTCTTTTTT CATATGAGG CACATAGCCT GCTCCCAATT CAGATATGTC 660  
TTCCATTGTC AGCAGGCTG ACCACCTTCA GCAGAGTCC TCCAAGATG CCAACTCCC 720  
CTTCCACAG TACACAGCC TGTAGTTGTT GTCTGCAAT CCTTGTATT TACCTCATTC 780  
TTTCCCATCT AGTCTCTCAC TGAGTTTTHA AGTTAGGGCT GGAAGAGCTA TGCCTTACTG 840  
GCAGAGCAG GAACCAATTT TTTTCTGAGG GAGAAGACAT TCACCTTCAC TATATGCTG 900  
GCAGGGCAC AGTGACAAA ACAAGATCA GCTTTCATTC AAGTTCCAGG TTTTCTTCTC 960  
TCCCTGAATG ATTACTGCA AGGTATATG AAGTAAGAGT TCCCTGTTCG ACATGACCA 1020  
TCCATAAGGG ATACTATATC GTTTTGCAAT GTTCCGCCA TTCTCCACAT TGTCTATCT 1080  
TAGTCCAGG CCGTTTTCAC TCTCAAAAA AAAAAAAAAA TATTTTTTTC AGCACTGGTG 1140  
TTCAAAAGCA AGCTTTTAT GCTTAATGCT TTACAGCAAA CTGTGAGAT TTCAGATTGA 1200  
GTCTTAAAA TTGCCATCA TTATCTAGCA GCATGACAG ATGATTAGGA GCAGTCAAAAT 1260  
CCTCTGAAT CTTTCCTTAA TAGGCAGCCA TTGTGAGAACT GCATGAGCTG ACATCACTAA 1320  
ACATTTATCA GCTAAGCCA AAACCAATA AAGGCCCAAG CCAACATCTT GGTCTCTTAA 1380  
AACCTGTCCA AATCTNTTAA GTCAAGCCA GTAAATCCAG GACTGTGCAAT CATGTCACTG 1440  
CAGCTGACAA TGATTAACAA TAGAGACAT GCAAACCCCA TTAAAGTTAA AAGTCCAAAA 1500  
CTAGTCAAC GCATCTCTTT ATTGGGAAA AGTGAGACTA TTATGCATTC TTGTAGGTT 1560  
TGCACCTTG CATGAGAGC ACCATTGCA TTTCTTTTCAT CTTTCAAGAA GCACCGGTAT 1620  
CTGTTCCAG GGCCTAACAG TACAAAAATA CATCTGCGCA TCACACTCTT GAACCCAAAG 1680  
CTGTTCTCAT TAAAAATPAT TTGGTTTGT AACAAATTA TGAATATCAA TCCAGGACCC 1740  
TGGTATAGC ATTATTAAGT AAACCCCTTA ATTCCAGCT TTTTGTGTTT TTTAAAAAAA 1800  
CCCACTCAC TAAGATTCAC AATTCATTC TACATACAAA TTAAAGCTAG TAAGAACACA 1860  
CTAAGTCAAC AAGTTTCTCA TTCTAAGATG CAAAGCCCTA ATCATCTGAA AGTGACAGG 1920  
GTAA 1924

(2) INFORMATION FOR SEQ ID NO: 47:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 475 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

60 TCGTGTGGGG CCCGAAAGAC AGGAGACAG TGAAGAAAC CCCAGAGACT TGTATTCGGC  
120 AGGAAAGCA TTGCGAATTC TGAAGCCCTG AAGGCGAAG AGGAAAGAG TGTATACAGA  
180 GCGCAATAG AACCTGCTTC ATGAAAGAGG GCGCACTTGG TAAAGAGACT CATTACTAAC  
240 AGAACTGTGG TCGCAATTTG CTGTGTCTTC TCTTTGAGAA AACCAACAG ATACACTTGC  
300 TCGAGACCA GGTGGGACA GAGAAAGGTG GAGAGAGAT CTGGAGAG AGATGAGAA  
360 TAAACAGAC AGCTTATTC ATTCTGTAA ATTCTATAT AAGAGCTCA GTGTGAGAA  
420 TAAAGTTATT CTACTAGTGG GAGTTAAGT GTTCTGTATT GTTCTGCTTT CCGTTACCA  
475 TAAATTAACCT CCCTTGGA CTACAGAGT ATGTCTTCC TTGACATAT GTGAA

20 (2) INFORMATION FOR SEQ ID NO: 48:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 346 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

60 AAGGAGAGA GACTGGATT CAGATTCAT TTATCAATGA AGACCCGAT GCGAAAGATC  
120 ATGTCTGAAA TTCTGAGCTT ACTCTCTCC CTCTGGAGC CTGCTGTGA TGAAGAGAG  
180 GAGGAAAGG ACTATACAGA GAGCGAATG AAGTCATCC ATGATTTCC TGAACCTGC  
240 CAGCTAGAG ATTAACTTT GAGCTGAC GTTAGAGACT GTCCAGATGG CTACTTGTG  
300 GCGAGATGA AGACCAAGC CAGAGCAGG CCCGAGCTT GCTAAAGAG CAGAGATCTT  
346 GCGGAGCCA CTCTGTGAA ATCTGCTTC CTCGACAGC TGAACC

45 (2) INFORMATION FOR SEQ ID NO: 49:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1366 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

60 TAGAGTACG CCCGACCCC CCCCCCAAT GCAAGTTAC TGGCAGGT AGTGGCAGC  
120 TTCTAACCA GCTGATTTT CAGCTCTCT GGAAGCTAC TCAAGATGA TACCCCCCA  
180 GTAAAGAGC CAGGTGATCT TAACTCTTT CAAGAGCAG GCTCTCTGG GAAAGCATG

240 CTCGACAGG CAGAGCAGC CTCGTGAAA TGGATACAA ACTGACTCT CAGCGAGCA  
300 GCGAGACTT CTATTTTAC AGTACGACA GTATATGTGG ATCATACAT CAGCTGGATT  
360 TTTGGTTTGG TGAATTTAGG TGTCTGTAC TAAAGGTATT TGAAGTCTCC TTGCCCCGTG  
420 AATGCTGAA CTCGCCAGC CGAATTGCTT CAGTTTCTCT GAGCTGATG TGTCTCTGG  
480 TGTGGGACA GGCCTTGCA TGGAGAGGA GCTCTGGG GCGAGACA GCTGGGGGTG  
540 CTGAGCTTAG GCGAATGAA GTTATTTAG GAGTGGTTAT TGAATTTAT GAGGTATTC  
600 TTCAATGCCA GCGAAGATG GCGAGATGA GCGAGTGTCT GCGAGTCTTC TGTCTGAGC  
660 TTTGTTCTCT GTGCCAGGG ACCGTGAAA AAGTGTGAG GCGCGTCA TCAGAGAGC  
720 TGTCTGTCTG CTCCTCCAG CAGTGTCTTG GGGTGTGATT CCCGACAGCT AGATGTGAA  
780 GCGCTTACTT TTCTGCCAC AAGGAGTGG CAGCGAGCTT AGCTGTGACT TGCACATG  
840 ACAAATTTCA CTAGCAGCT CTAGGCAAG ACTGGGCAAT TGAAGAGAG AGACGAGCT  
900 GTAGATCTGA CTCGAGGCG GTCCTTTCA CTCCTGCTG GCTGTCTCT GTCTTAAGT  
960 TTGTCAAGT TGTCTTTTGT TGAATCTTC AACATGATGA TAAAGACTGG AGGGAGATGA  
1020 CCCGCTTGG AGGTGTTCT TTAAGCTCAT CAGATTAATA GGGCGTGGG ATGGTTTGG  
1080 AAGTTAAGCA GATGATGAT GTTTTAAAG CAGAGCTTGG GACAGAGGCT CTACAGCTA  
1140 ATTCTCTTC CTGTAGACTG AACAAAGTTC TAAATTAAGT TAAAGAAAG AACAGCTGTC  
1200 GTACAGCAA GTTCTGAGG CAGATCTTTC AGTTTCCCTT GTTACAAAT GCTCTCAGAT  
1260 TGTCTTAAA GCAAGAGCC TTAGGGGGGG TGGCAGACA AACAGCAAT GTTAGAGCTG  
1320 AACTGTGAC AACAGGGGAT TGTGTGTG TGTCTTTCT CTCTGTGCG AACTTTCTCA  
1366 ATTAAGCTTA TTTCTTATT ATTAAAAAAA AAAAAAAA AAAAAA

45 (2) INFORMATION FOR SEQ ID NO: 50:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

60 GAGTAATTC CTGTACCA CTGATCCAC CAAGAACTT TTAATTTTC CTCGAAATTC  
120 AATATTTTA CTTCTGTTAC AAAGGAAATT TTGCGATAG CTCTTGGCT CCCAGCTTC  
180 CATTATAGC TAAATGATTC CATTCTAGG CCCAGGCTTC TCTTACAGA TGTGTCTTCC



5 TTATACCTT GCATGTATA ATAGATTTT GGTGATGAA AGAACCGAA TAGGCAGAT 240  
AGTCCGCCCA GGCCTGATA TCCATAAAG GCTTGGAAT GCATATGTA ATGTGCTTA 300  
GTCTTTTGT TGTTTTGA AAAAAACA AGATGGGCTC AGATGATGC CTAGGTAAA 360  
ATGCTCTTA GCTGTGACT CATACTTTT CTTTGAAATG AGTAGTAAA GGAAGGAGA 420  
GGAAGGAAA TTAAATGTC TTCTAGTAT CTCTGACTC AAGTCTGACA TAGAGATAA 480  
TAACCTATAT TGAATGCCA AGAATGTAT CTGAACCAAG AGAACGTTT GACACATTA 540  
TCATGCTTTC ATATTACATA TTAATGTAAA CCAATTATA AICATATGAA ATATCATTTG 600  
15 CACAAAGCAA AGCAACTTA ACCTTTGTT TCTTTTCTA CATAGCAGAA ATTGATTTT 660  
TTTTATTTT TTAGGGGAA CCTATATAT TATGACCCGAG TATGTCTTT TGGTGACTTA 720  
ACCTATGAA TTCAGGTTAC AATGTAGTTC ATTCTAGATG GTTACTACT TCGAAAGGAT 780  
GTTGTGCTT TATGTGACAC GAGCCAGAGC CTGCTGGGGA ATAAACAAG CAGGTTTCAT 840  
GCCAACCA ACTGTAGCT TTAGTGGGCA GATGGGAGT GGTTCACAGA CTTCCCAAAA 900  
25 TGTGGGGCT TTGGATTTT CCHACACATC CCAGTGTGT TGTTCATCT TCTCTTTTC 960  
ACACTCTTGG ATGGATWAT TGBAATGAT YKBAATWY YKBAATWY CCAATAGCC 1020  
WTGRCGACC ATTCTWATG ACACATAC CAATATGTC ONTATGTG AAATATAGA 1080  
AACTGTAC CAGCTYKSA KYACCONHA WTTTTCCTT GTTTGGGA TTGATATGA 1140  
AATAGCAGG CTAGGAAT ACTGCAAGT TTTAGCTGT GGTATATACC TTAGGTTAT 1200  
35 TTAATATTT GTATTTTAT TTAAATGTC ATGAATGTT GAAGGACA AAATATCAG 1260  
GGATGCTCT TTGCTGCGG TCTATTTTC ACCCTTTT CTGTAGGAA AAGACAT 1320  
40 GTCTTAATCT ATTTTAAAG TTTTGGTAT AGTTCTAAT TCCATTTTA ATAAAGTTT 1380  
TWTATAAAA AAAAAAAAAA AAAA 1405

(2) INFORMATION FOR SEQ ID NO: 51:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 504 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

55 CGGATTTTCT AGGACCCGAA AAAAAAAAAA AGGAAAAA AAACCCNCA AACCCNCA 60  
AAACCCNCA AAAAAAAAAA TCCAAAAA CAAAAAAACT ATAAAAAGA AAGATATA 120  
60 AACTTCAGA GAATTAAT TTAATTAAT AACTTACGA TTAATTAAT AATATATAT 180

5 TCCTCTAGCA ACATATCTT GCGTCTCTC CTGCTCTCAT AATGAGACA TAGCGATTC 240  
TCTGCCGCGG CCCCCTGCTG ATGCTCTCTC GGGTCTGCTT CGGGCTGCGG TCTCTGGGA 300  
CCCTCCAGAG GTGAGAGTGG GCTGATGCC TGGTCTCTG GTGGTTGATG GTTTTCTCC 360  
CCCTACCTTT TTTTCTGAG TTATCTGTA TTGATTTT TTCTCTGTTT CTGATTAAC 420  
10 CACCTCTGG GCACAGGATA ATAAACATG TAATATTTT AAGAAGAAA AAAAAAAAAA 480  
AAAAAAGTGG GGGGGGGCCC CGAA 504

(2) INFORMATION FOR SEQ ID NO: 52:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

25 NAATATCTT GGCAGTTTA TTACAGAGA CGATTAATGA TTCAATGTG ATAGGACATA 60  
ACNTACAGAG AATGAGACTA TCCAGGAAT GGGAGAGGC ATTGGAACA ACATGATAT 120  
CTCAGGACA GATGATTTCA TTCTCTATT GGTAGGCTG GAGCANGG TCAGAAATAG 180  
CAAAAAAGG ATACAAAAG CATATTTGT CACCCAGCT AGTGGANTA GCTGGCCGAG 240  
35 TAGGGAAT GCAAGTTTGT CTCTACACTA AGTTCTCCA CTCTTGATA GCCTCCAAA 300  
ACAAATGTTA GGGGAAAAA AGCCAGCTGG TTATGAAAAG ATATATCTCA TTTCATTA 360  
AATCAATGT CAATGCTGT AATGAATCC TTTTATCTC AGGACAGAG CAATGCCCTA 420  
40 AACAAACC ACCTCAGAG CCTGTATGC CAACCTAGAG GGTACCGAAA CACAACTTA 480  
GCATAGAGT AAGATCTCT ATGTCTTTTG GTGAGGCAA AGCCATTTG TTGTACTTC 540  
45 AAGGAACAT CTTTCTACA AGCTCTCATC ATATGTTATG TGGCAGAGT CTCAGTTGT 600  
TTGCACCAT GTGCTATAG TGAATATCG CTGAAGGTT AGTTTATC CTGAAACCT 660  
70 ATTACATTT GCCAGCTGAT GTCCCTGCTG CCACTTAAA AAGGCTTGG TCTGGCATG 720  
50 GCGAAGAGC CTGTGTCTC CTGTGCGCA TTCTNGGCTC GAGGCCAAT NOCTTAT 777

(2) INFORMATION FOR SEQ ID NO: 53:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 602 base pairs  
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

5 ATGACATGAG TGTATACCC TCGATCTT GCAAGTGGC ATGAAACAT GCTGTATCA 60  
CTCTGTGAC GGTATTAATC CATTATCCA CAAAAGACA CATTCATCA TCAACATTA 120  
CATGTGTGG GATGACGAG TCAATGTT TGAAGGGAG TTGTTCCTT TTTTTCCT 180  
CATTAATCT TTAAATGTT GTGAGTATC AAACAAACA ACAAAAAA TGTGTGAAA 240  
AACCTTCAT AGCCTTTTC TATCAAGTC TTAAAAATG AGCTAAATA CACACATCT 300  
GCAATTTTT TCTAACATG ACAGATCTT TACCTGCCAT TTAAATTTAG CCTGTATTT 360  
TTCTACGTA TTATTACATG TGAATGTAT TTGTATTTTA AACAGAAA AAAACATCA 420  
AAAAAGAAA AATTAACTGT AGCCTTCAT TATTAATTA TATTATTAAT ATTATTTGA 480  
CAATTTGGA TACTGTGAA GTTTATCTC TTGCATATAC TTAAATGGA AGTATTTACC 540  
CTTAAAAAT CAAAAATTA TTTTCAAGG TTCCGTTTT GTGTGTGAA AGAGTAAAT 600  
GA 602

30 (2) INFORMATION FOR SEQ ID NO: 54:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1749 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

AGTCACTGAC TTGAGGCCG TCGGGGGAG TTCCGCGCG AGAGCTGTT GATGGGAATG 60  
CCTCACTTCA GTTTGAAAG GATCCGATC CAAGGGGTT AAAAGACAG AACCCGATC 120  
CCGACACAA CTTCGCGCTT CCTTAAAGG CACACCGCC TACGCAATGG CACCCGAC 180  
CACTGTGTA AAGTGTGTT GATGGGGAG GCGGCAATGG GCAAGATCT GCTGTGTAG 240  
GATTATTCG AGAGACCTT CAGCAAAAC TACAGTCCA CGGTGGAGT GGAATTTCT 300  
CTGAAGTTC TCAATGTTG TGAATGAGG AATAGTGGGC TTCACTGTG GGAATTTGCA 360  
GGGCAAGAG GCTTCACCTC TATGACAGA TTGTATTAAT GAGATGCTTC TCCCTGTCT 420  
AATATGTTG AGCTTACAA TGCACATAC TTTCACACA GCGAGAGTG GAAACAGAC 480  
CTAGACAGA AGCTCAACT AACCAATGA GAGCCGCTG CCTCCCTGCT GTTGGCAAC 540  
AAGTGTATC TGTCCCTTG GCGAGTACG GGGACACAG TTGACGCTT CAGTAAAGG 600

ACGGTTTCA CAGTTGGAC AAAAACTCA GTCAAGAGA ACAAATTAAT TAATGAGCT 660  
ATGAGATCC TCAATTAAA GATGATGAG AATTCAGAG AAGATATCAT GTCTTTGCC 720

5 ACCCAAGGG ACTAGTCAA TGTCAAAAC AGTCTTCCA GCTGTGCTG CTGTAGTAG 780

TGTGTGCTT AATTTCATC CAGTTCTG GAGTCTTTT AAGTCTCTTC CTTTGTGTT 840

10 CCGACTGAC CAATTATTA AGTAAATGG AATTGTCC TGAATCTGT CCGATAGGA 900

GGCCCATGG TCACTTAAA AAGACATCG GACCCGATG GAAATTTCCG ATCTCTGGA 960

15 TTAGCTTTC ACATGTGCT GTCTCAATG AGTCCGATG AGTCCCTGG GTGTAAATC 1020

TTCTGATGAG CCGTCATTT GTATCCGGA AATTGTGAG AAGATTTGA AATCAGACC 1080

20 TCGTTTTAG AGATCATTA TCTCACATC TTCTGAGCT AATTTCATG TTGATATTTCA 1140

TTGATATCAT GACTTCATG TGAAGGAAA ATGAGATCA ATGTCAATTC CCAATTTCT 1200

25 TGTAGCCCT TGTTCAGAT TCTTCTGTC TTGAAATGA AACATGAT TCTGGAATCC 1260

AGAAAGAGA GTCTGGCAT CTGTGATTT TTGCATACA GAAATGTCCC AAGATCTACT 1320

30 GTATTTTGA AACTTCATC GTCAATATA AGTTCCTCT GTCTTGCAT CAAGATTAAT 1380

CAAGTTTTT GCGCGGCAT GATGCTCAT GCGATTAATC CGAACCTTG GGAAGGCCAA 1440

25 GCGAGGCCA TGCATAGAG CGAGAAATC GAGACCAAC TGTGTAGCAT GCGAAAACCC 1500

30 GATCTTACT AAAATGACA AATTAGCCA GCGGTATGG CAATGTCTG TAATCCGAC 1560

35 TACTTGAG AGTAAATGG GAAATGCTT TGAATCTGG AGCCAGAT TCGATGTAAC 1620

CGAGATCAT CGACCCACT TGAAGCTGG TGAACAGAA GAACTCTGTC TCAAAAAA 1680

40 AAAAAAAA AAAATCGAG GGGGGGCGG GTACCCAAAT CCGCTGATA GTGATCTTA 1740

45 ACAATGAAA 1749

(2) INFORMATION FOR SEQ ID NO: 55:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1896 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

55 AAGAGATGG GCTCTTAAT TTCTGAAA AACAAATGG AGTACTCAT TTTCATTA 60  
CATTAATTT CTTCAGTGA ACTACATAT GTTCATTAAT GCTTCATGAG GACTCATCC 120  
CCTCTGTCT ACTGCTCCA AATGACAT GTACAGTTCA CCGCTGCTCT TTGTCTAAT 180  
60 GGTGCGCTG TGTATATGG AAAATGACA GGTGTGTAG GGTGGAGAC ACAATATTT 240

5 TTTATAGTCT AGACCTTTA AAAACCCAG CAGATGTAA TTCAGTATTT GTTATTCGC 300  
TTTTTTTGA CAGATGTTTG AATTAAATG AATTGAAGG GAACTCAGA GTACTAGAC 360  
GTTTATTAAA AGAAAAAAA TGTCTTCCA TGTCTGTAA TCACAGAGG AGAAATAC 420  
TTGTTTCCTT GATCTGTAG AGCTCAGAT AACCTGGGC GAGCTGTAT TATTTATAT 480  
10 AATAAGTAG TAGGAGTTA ATACTGTTT CTCTGTGTC CAGCAGAT ATTACAATT 540  
CTTTGAACC GTAAATATCA GAATGAATCC TCTTCCAGG GAAATGAACA GAAGTTAAT 600  
GTTTACAGT GTTTGAATTT GTGATCTGAA ATACACAAA ATTAAACAA TGATTTCTCT 660  
15 AATTTTCCA CTAGAGGAG AGAACTTGT GGAAGATTC TTTTTTTTC TTTTTTTTT 720  
CTTAAGAAG GGCAGCAGG GTAGTAACTT AAAATAGTG CCCAGGCATA TGAGATTTGT 780  
20 CCTACAGGT TAAAGAAC ACCTGTCAC TGTATGGCTT TGCCCTGAG TGGCCAGGGA 840  
GCTCACTTG ACCCTGCCAT GTTGGTTTGA CTTACTAAGA CACAGATTC ATTGTTTCC 900  
TGACAGGG TCTCAGACC TCGAGGATG TTAAGTAAGA GAAGAACTT CTTTCTGAA 960  
25 TATTGACATG TAAAGACAA AAGTAATTT TCTGAATTC TGCATTTG AGAATCTCC 1020  
AAGCAATTGA CAGTGATTTT AGCTGTTTC AGCAATTTTC CATGAGACT TTCATACAT 1080  
30 TGACTCTTTA GTTCAGAGT TCCCATGAT TGTGAGCAG AATATTATCT CTTTAGCCCT 1140  
TGGGATCCA GCTGAGACA ATCTCTTGA TTTTTTACC GATGATATTA CAGATATCAT 1200  
TCTGTGTGA TGCATGACT TGAAGAGTT TGGGAGCTC TTTAGCATA TCAGCTTAAA 1260  
35 GCAATGAAA TCACAGTGA TAGCAGTTCT CATTCAGTAA TTTCTTACAA GCAGCACCC 1320  
AAGGAAATA TAGTCTTAAT CTTTACTAT CACTTCTAAA TTTAATGTGA ATTTCATACA 1380  
40 TGTATATAGT TGTTTTCTTT ATAAATTTAT AAAAATTTAT CATCGGAGT TTAATCTCCA 1440  
CTTCCATGCT ATCGATGTG TTGGCTCCA TGCAGAACT TCGAAGAAA ACAGGCAGA 1500  
45 ATCATTTGC ATAAATGACC AGATCATCAT TTCTGCAAC TGAGAAATAT ATTTCATCAT 1560  
TGTCTCTAGA AGCTGCAT TCTTTACTTT TCTTTGTGTC ATTATATCT AGCTGCATC 1620  
ACTGGATAAT GTGGATGAC TAGAGAGTC AVATATCACT GTAGGTACA GTTAGGGSTA 1680  
50 ACACITTTGA GGTATATAT TTTTAAAAA CTTTCTTGA ACTCTGGGC CAACATGGT 1740  
GAAACCCGT CTTCTACTT AAAATACCC AAATTAAGC CAGGGGCTG GATGGGTGG 1800  
GTGCTGTGTA ATCTTCAGT ACTTNGGGA GGGCTTGAG CAGGGAGA ACTGCCCTG 1860  
55 AKCCCCGGG NCGGCCAGNA GGTITGCCAG TTGACT 1896

## (2) INFORMATION FOR SEQ ID NO: 56:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1753 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

10 TCTTTTAAA ATGACATTTT GTGGGCTCA CACATATAT GAATAGTAC CCTCTAAAA 60  
AGAGAAAAA AAAATCAGGC GGTCAACTT AGAGCAACAT TGTCTTATTA AAGCATAGT 120  
15 TATTTCACTA GAAAAATTT AATATCAAGS ACTATTACAT ACTTCATPAC TAGGAGTTT 180  
TTTTTAAAT GACACTTAAA ACATCTACTG AAACTTTCAT CCAATCACA CCTGTTTAT 240  
TTTCTTAAA CATCTGGAA GCTTAAGCTT CTGAGATCA TGTGGCAAGT GTGATGGCCA 300  
GTAAATATCC AGAGAGATG TTTAGTAGCA ATTAAGGCT GTTTCACCT TTAAGACCA 360  
40 GCTGGGCTGT AATGATTTCT GGGGCCAGAG TGGCATATG TTTTACAAA ATATATCAT 420  
25 ATGTCAATG TTTGCAATG TGTGCTGCTG TTGAATTTT GAAGAGCCAG TTGACCAATC 480  
ATGAGAAATA TTAATTTCTT TCATATGGTT TTTGGTTGAC TGGCTTAAAG GGTTTCTCAG 540  
AATATCTATG GCGAGAGCAG CATACAGTTT TCCATCTTAA TAGAATGAA ATTAAATTTG 600  
TATCTACTGA TAAACAGATC TGGTTCACAT GAAAAAATAT CATTTTATCC GTCTTTTAG 660  
TATATGTTTA AATTAATTAAT TTATGTGTCT GCATATTGCA GAAAGCTCT GAGAGACACA 720  
35 GTTTCGCAT ACTCTTTCT GACCAATAGT GCTGGAGCCG TTGCTTCTC TTGGGAGAA 780  
GGAAGGGGTG TGTGACATG GCTAACATC TTCAAATACC CAATTTGTGA TAGCATTAAT 840  
40 AAAGTATTTA TTTTATGCTT CAGTATATTA TTATTTAAT TTTTAGSTAA TGCCTATCT 900  
TTGCTCTATT AAGGAAGAA GCAATCAGTA GAGATTCAG GATAGTTTGT TTTAAATTTCT 960  
TCCAGATTAAC ATGTTTTHAC AGTGGCTGCG TATTAGGGA AGTATTTCTT CVATACAACT 1020  
45 TGTTTTAAAC TTTGAGACA TTAGCAGAA TTATGCAATG GTTTGTGAG ATACGAGCTT 1080  
GATGGTCTG TTTAATCAGT TTGCTTCCA AGTGGCTTAC TGAAGAGCC CTAGAGCTGG 1140  
TAGAATTAAT AAGGATTTCA AAACTTTCTT ATCTCTTCT TAACTTACC AGCAACTAG 1200  
50 GATTTGATA GCATGAATG GTATGATGAA GAAATTTGA CCAATTTGT TTTTTCGTG 1260  
TTGTTGTGT TTTGAATTTG AAATCATCTT TATTCCTTTT AAGAAATGTT ATGTATGAGT 1320  
55 GTGAAGATCC TAGGCAACTT ATGCTCAGAT ATTCATGTA AGTCTCCTT CACCTGTATC 1380  
AGAGTTTCTAG ATGCTCACT GATAGTATGT ATTTCTTTAG TAAGAAATG TTAATAATAC 1440  
AATGATCTTT TAAAGAGTG ATCAGTTCT GTATTTATG TGCATGCTCT GGTCTTAAGT 1500

GGAGCCAAAT AAACAAGTTT GATATGATTT TTTCAGATGT TGAATCTGAC ACACATGACT 1560  
TTGAAATATTT CCTTCAGATCC TGATATAGCA ATTAAGAGAG CCATATATAT TTCTCTCTTA 1620  
TCTTGAAT TTCACTACT TTATGTATTA AGTTGTAT ATTTGTATTA ATCTGTAAA 1680  
GAATAAAAAG TGAATTTAAA TTAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1740  
AAAAAAG GCG 1753

## (2) INFORMATION FOR SEQ ID NO: 57:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1220 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CGGAGATTTA CTCAGCCCC GGTGTGTGTC TGTGGGAGAG GAGAAAGAT TTGTAAACC 60  
CGAGAGAG TTCTGCTTAC CCGAGCCCC TGTGTGTGAG AGACCCCCCG GTGAGCCAC 120  
GCTCATGAG TGTACAGAG AGCGAAACC TTCACTGAG GACTGTGGAG ATAAAGAGGA 180  
AGGTAAAT ATTAACTCA AGGTATGAG ACAGATGAC AGTGAATTC ACTTCAAAT 240  
GAAATGACA ACACATCTCA AGAACTCA AGATCATGAC TGTCAAGAC AGGTGTGTC 300  
AATGATTTA CTCAGGTTTC TCTTTGAGAG TCGAGAAAT GTGTATATC ATATCTGAAA 360  
AGAACTGGA ATCGAGGAG AGATGTAT TGAAGTTTAT GAGAAACAAA CCGGGGGTCA 420  
TTCAAGAT TTGAATATCT TTTATTTT TTCTTTTC CTCATGCTT TTTTATTTT 480  
AAAAATAGT GTTTGTAT TTGTGTTC AAGCGAAT GAAACCTGC ACCCATGTC 540  
TTTGAATAT GTGTATTT GAAATCTAG GTCATATAT CATTAATGTT TGTTTCATTT 600  
GTGTGATTT TTGTGTATCA AGCTTCATC CCTTCATAT TTACCTGTC TTTTAAAA 660  
TTAGGTGTC ACAGAGAGT GACCTTTTC AGACATGTC ATTTTCAGGC TTGTGTGAT 720  
AAATAGAT GACAAATGCA AGTGTTCATA ATGATTTTC AATTGGCCT GATGTCTAG 780  
CATGTGATTA CTCACATGCT GAGCTGTGAC TTTCAGTGGG AGATGAGAT TTTCAGAGA 840  
ACTGAACTGT GAAAAATGA CCTTCTCTA ACTTGAACT ACTTTTAAA TTTCAGGTC 900  
TGACAGAAA GAAAGAGAT ATCAGTTGA AGTCAGAG AGATGAGAG TGAAGTAT 960  
GACTACTCC AAGATGGCT TCACTGAGA AAGCGATTT TGAATTTT TAAAACTT 1020  
GTGAGAGAT CCGAGAAAG TTCTAATTT CATTAAGAT TAAATAGCT ATACATGAG 1080  
AATGATATC AACAGAGAC TGTCTTTT GATTAATTT GATCTTTG GCTGTGATA 1140

TGAGTTTAAA ATGACATGAG TGTATGAG CTCATTTAAA ATAAACATA TTGTAAAAA 1200  
TCAAAAAA AAAAAAAA 1220

## (2) INFORMATION FOR SEQ ID NO: 58:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1049 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

TGTGCTGTC AGACAGACA TCTACTGAG GTGGGTGACC TGTGTAGCA TCACTGACTG 60  
GAGCTGCCC TGAATTTCTG GTTCAGGCCA TCAAGGAGCC AGAGCGCTTG GAGAGATGT 120  
GGGATCTTG GCGAATGGG AATGTGTCA GAGAGAGAG CCGGAGTTGA GACACACTAC 180  
CGAGCCACA AAGATGACA TTCTTGACA GACTCTCTC AATAGAGGCC ATGTGCTCC 240  
CCAGAGGAT CTGAGCCCC GCGACAGACA GCGAGTGC AGCTGTGGTG GTCTGAGTC 300  
CCCTTCAT GACTGACAC GCGAGCTGAT GAAATGGCC TTTCGGCAT GCGATCTGAG 360  
GAGCATGCT GTGAGAGCCG TGACCTGCA CTGTCTCTC TTCTGCTCA TGAATCTTG 420  
TGTGTGACC CTCCTCTGAG TTGGCTTGT GTACGTGATG TCCAGCTGA GTTCAGGATG 480  
ACCTGAGAG GCTGATGAGG GTGGTTTTGT TGAAGAGAC TTCTGAGCC TTGGTGTGAG 540  
AGAGAGATA TTGAGAGGG ATCTGTGAT GCTTTGAGG TATGATGACA GAGGAGACA 600  
GAGGTGTG TTGCCCCCTT GTTTAGAG TGAGGAGAG GAGAGCTTGA GTCCAGCAT 660  
TTCCAAAGT GTGGTGGAT CCTTGGTTC CCGAATACT TTGAGTGTAT ATGGGAGCTG 720  
CATTAAGTG CACAAATCA GAGCAGAAA GCGATGCCCT TCCCATTTCT CTCATGCTT 780  
TTATCCGAG AAGATGTGAG CTCATGACA AATGTTCAG AGCTGTGAG AAGCATGCC 840  
GAGCTGCTT CTCCTGAGG ACAGAGACT GGGCTTGAAG CAGTGTGAAA AAATTCAGCA 900  
CTTTTTCAG ACTGTTTTG TTGTATAGT AATTTTAT TGCATCTTT ATTTGTGAG 960  
ACAAGTGTGA GTGCATCTT ATTTATGTG ACCTTTTCA TAAATAGAT TAAATAAAA 1020  
AAAAAAA AAATCTGAG GGGGGGCC 1049

## (2) INFORMATION FOR SEQ ID NO: 59:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1776 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAGAGGATG TGGAGCTAGA GGTCCCGCAT GGTCTGTCTGG ATGGGAGCA CAAGGCTGAG 60  
10 GGACTGGGATT GTTAAAGGCAC TAACTGCTTC TGGGGTGAGA ATCAGACATG GGGGACCTCT 120  
AGCTTCACAT CCTCTTCTCT TCCAGSCTCG GACATCTCTA GCCAAGTCC CCCACACTCA 180  
15 GTCCAGTCAAT GAGTCCGGAA GTGAAGTGA CAGGGCAGAA CCAGGAGCAA TTTTCTCTCC 240  
TAGCCAGTTC GGGCCAGGGG GCAAGGCTGG CCACACTCAT CCATCAGTGG CTGGAGGGCC 300  
CTGGTGTCTA CTTCTTTTGA GAATCTCTGG ACATGCCCAA TGTTAGAGAG CTGGCTGAGA 360  
20 GTGACTTTTC CTCTACCTTC CCGCTGCTCA CAGTGTCTTC TTATGGGACA TAGGCTGACT 420  
ACTTAGCTGA AGCCCGGAAT CTTCTTCAC TACAGAGGC TCAGAGGAT AGCTTTGAC 480  
25 ACCTTCAGT TGTCAACCTG GCTCTTAAG TAAAGTGTAT CCAATATCCA GTTCTCTGG 540  
AGGCTCTTCC CTTGCTTAAT GTGGGAGC TGGAGGACT TGTGATTGAG GCTGTGTATG 600  
CTGAGTGTCT TGTGTGCTTC CTGAGCAGC GCAACAGCG GGTGAGGTT GACTACAGCA 660  
30 TGGGGCGGGA CATTGAGGC CAGAGCTCA GTGCCATTGC CCGAACCTC ANNNANNC 720  
ATTAAAGTTA CGAGGAGC AGCAGCGCA GCCACATCTC AGGAGCCCTGA GCACACCTG 780  
ACTGAGCTGA GGGAGCAGC TCTTGGACC AACCAAGGCC AGCCAGCAG AAGGCTTCAA 840  
35 AGGCGAGGG CTTCCAGGG ANGCCAGGA TTTGTCTCAA GTGGATTGA AAGTACTGTC 900  
GTTTCTCTCC TGGGATCTG GAGTCCGAGC TCCCTGCTCG CTTCTTAGCA GTCTCTAGAG 960  
40 AGCTTCTCTT GCTCCCTGGC AGCTGTAAT CCTAGGTTCA TGAGCTTCA CTTCCCTTAA 1020  
CCCAAGACAT AATCAGAC CTTCTTAGG AGGAGKCAA TGTAGTCTAT GTTTTCTTGT 1080  
45 GTACTTTCTG TTTTCTTGA CTTCAATGT TCCATTGCTC CCGGCTCCA TGTCTCTGCC 1140  
CTTGTCTCTT TAAGAGCTCA GCATCTGTCC CTGTTCATTA CATGTCAATG AGTAGGTGG 1200  
TAGCCCTGAT GGGGTGTCT CTGTCTGAG CATACCCAC AGGGTTTTTT TCTGCCAACC 1260  
50 CATCCCTGCA TGCCTGATCC CAGTTCTCTA TACCTTACCC CTGACTTAT GAGCAGCTC 1320  
TGAAGAGCA TAGGGCCCCC ACCTTTACT ACACCTGAG AATCTGGGA GCAATCTCTC 1380  
CATGCCAGG GTACCTGAC ATGTTCATCC TAGAATCTG TCACACTACA GTCTTTCTT 1440  
55 TTCTCTCTC TGGCCCTTGG GTCTGAGGA TCTCTCTGCT TCACCCCTAG AGCTTAGAA 1500  
TGGAGCGGT TTTCTTAAT GTTGAGAT GATTTCTTCT TGGCCCTGAG CATCTCGGA 1560  
60 AGCTTGTATG CAATCTGGA AGGGTTAAT CTTCTTTTGT GAGTTGTG GGGAGGGAA 1620

GGCTATATAG ATTGTATTAA AAAAAAAAAA GTATATATCC ATATATCTAT ATATATATATG 1680  
1740 ACCAGAAAT AATCTATGA GAATCTATC TACAAAGAAA AAAAAAAAAA AAAAAAAAAA  
5 AGGAATGGA TTTGAGCTT ATGATACCG TCNACC 1776

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 443 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ACAGTAAAT AATAATATA TAAATTAAT TAAATAAAA ATCTGAGCTA ATCTGAATTA 60  
120 ATTGAGGAT TTACATGAA ACCGAGGAT TCTGGCTTCC CAGGACAGT CAGAGAGCT  
25 AGTAGGAC ACTGCTGTC TTGGCTACT TCTTTGGAC AACTGGAAT CTAGCTCCT  
TTTTTTTTT TTTTGGGCC ACTTCATCA TTCAATGAC CTGGCTGGCC TCTGCAGGTA 240  
30 AGTAGTATG CACAAAAAT GTAGCAGAG TTTTGTGCT GAACTAGCTG GTTTCAGCTC  
360 CAGCTCTGCC ACTGTGAGC ATGACCTGCT GCGGAATTC NCGAGGAAT TTTTTTTTTT  
420 TTTTTCATG CTCAGTCCC CCTATTGGAG AATCTGCCC CCCCCTGGGA CAGAAATGTC  
35 ACCCTGGGCC CCGAATCCC TGA 443

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2888 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTAATGTGT CATATACCAC CAGGCCAAC AGAATTTATA TGAGCTGGAT GAAGATGATG 60  
120 ATGTATAGC TTGGTCTCT ACTTAACAGA TGAAGTTTCC AGCCTCAGGC GNCCTTCTCC  
180 ACCAATGCC TGGGCTAAGC AGTTCCAGC TTTCATGTC CAAGGCCCTC CTTCTCACA  
240 AAGTGTTC GAATATGCA TACAGACTC CTGCTCTGCC TTCTCTTAT CGAACAAGAG  
300 CCTTGACAA CATGTCCCG AACTGTGTA ACNAGAGGA ACCGCCCAA GAGCTGGCAG  
360 CTGCTGACC TTTTCTCAGC CCATTTGGAG GCACCAAGAT GACTGTGAAT AATCTGCACC

	CTCCAGTCCAC TGAAGACAC ATTGTGACG TTTTCTGTGT GTGTGGGCG CTCAAGGAG	420
	CTCCAGTGTG CCAATCTGGG GTAGCGAGAG TGTGTGTTGT GAAAAAGAC GATGCCATCA	480
5	CGGATTTTAA GAGTTACAC AACCGGTGTC TCGACGGGCA GCGATGAGAG TCCAGCTTTC	540
	ACATGAAATG GATGTGATTC AACTGACAC AGCCCATCTT GGTGCGGCTG AGTGAACACG	600
10	CATCATGAA AAAAGAGACG GAGCTGCGTC GCAAGGTGAA CTCTGCGTCC TCTTCCACAC	660
	CGCTGTCTGA AGTGAACCTT GACACATGTC TGAAGGCACT GTTCAAGTCC TCAAGGGGCGT	720
	CTATGACAC GCAAGCGACA GAATTCAAAA TCAAGCTTTC AGCAGGGGAG TGAAGGACAC	780
15	AGAAATGAGG GCAAGAGAGG GTGGCTCTGT TTTCCCAAGG GAAAGCTTAT GACCAATGGG	840
	CAATCGACAT GAGAGACCTT GATGTGAGAA AGGATTGACA GGGATTAAGA GCTTCCCTAC	900
20	TGATGAGAC CGACCTTCTT GTGTGTGTGT CTGCGCTGTC CTCTCTCTC TACGTTTAAAG	960
	TTTCTGTGAG TATGTGTTCT CATCTACATG CCAAGGTGAG CTGTGTGTTT TCAAGTGTG	1020
	CGTCCCGAGG CCGACACCCC AAGCTGATTT GTTATCTGAA AATGTACAC TGAATGCTCT	1080
25	GAGTGGCTTT CTGTGCGCCC CATGGAATGC AGCGTGGGGG CTGTCTGAGG GACCTGCTTT	1140
	TTTCCAGAGG CCGAGGGGCT GCGTTTCTTT TGTGTGTGTT AACCTTTTCA AACAAATGAG	1200
30	GCGATGAGAA GCGCTGTGTG CTGACAGGGA GCGAGGTGCG CATTGAGACT GTGCGGCTCC	1260
	TCTGTCTGAT CATGTGAGGT GCGTGGGTGAG GAAACAGTTC TGAAGGCTCT TGTCTGTCTC	1320
35	CGAGTGGCTC GAGGCTTCAC TAACTGGGAG GCGCAAGTGA GCTTGCACCG CTGCGAAGAG	1380
	TCTATCTAAG TGTGTGGCTT GAGTGTGCGT GTGTGTCTCC CCGAAGAGAA GTTCTCCAAA	1440
	GTTCAGCTTT CCGTTTTCCT TGAATGTGTC TGAATGCCCC AACCCAGTTC TCTTTCCTTT	1500
40	CTGGGTGTCT TTGCTGGGAG GGGGCTGTGT TGTGACCCCT CGCGTTTCTC AACTGTGCTG	1560
	GCACTGACAC ACAAAGTGTG TTTGTGTGAC GCGCAGCTCT GTTCTGATG GCGCTTGTGA	1620
45	AGGCTGACGC TCCCATGTG GAGTGTGAG GTTGTGACT TATTTGAATG GAAAGATGCA	1680
	GTTTGTGCTT GAGCTTCAAT TTCTGAGCTC AGTTGTACAC AAGACAGTGG TCAAGGATTC	1740
	CTGTAGGCACT ATTTCCAGCT GCGACAGAGG GCGCACTGTT GTTCCCACTT AATGTAGTGA	1800
50	CGCGATCAAT CCAATGACAG AAGATTTATTT TCTGTGCTTG GCAAGAGAGG AAGATGTCAAG	1860
	GAGCAAGGCG AACTGTACCA GCGAAGGTGT TTTCCCAACA TAAAGCGAGA CATTTGAGAC	1920
55	GAAATTTGAG AAGCCAGGCA GTTCCGTAAAT GAGCAGGCGCA GTTGTGTAC TGAATGTGTC	1980
	CGTGTGTGTT GCGATGTGAG AAGATTCAGG CTGCGAAGAC TGAAGCCACT TGGGAGACAC	2040
	GTTTGTGAGG GCTGTCCAAA ATCAATGACA AGTGTGTGAA AAGGCAATG CCGAAGTAC	2100
60	TCAAGACTTC CAAAGATTTC CTGTAGGCTA GCGCAGTGA AAAACCAAG GATCTCATGTT	2160

	TCAAGGCTTC AGTGTGTGAG GCAAGAGAGA CCGGAGATTT GAAACCACT TCAATGTGCA	2220
	GACTGTGAGG CTGCGCAAGA CCGAAGATTC CAAAGAGAGG GCGTGTGTGT GCTTCACTTG	2280
5	CAGTTCTTAA AAGAACTGTC GTTTTGTATTC TCCTAAGCTT TTCAATGGG TCCAGACTTC	2340
	TCTGTACAG CTGGAAGACA TTTCTCCAC ACTTTTCCCT TCTGTGCGCA AAGAGATTC	2400
10	CAGAAAGCAG TAAAGACTG TTTTTCAGCT ACTGTGAGCG GGGGCTTCAC TCTTGCACAT	2460
	GTGCTTAAAG TAAAGATGCT AAAATATCTC CCGTCAATGC TTATATCTCC CTGTCAATCC	2520
	AAAGCAATTA TCTGTGTGTT GTCAAGAGAT TTCAATGAGT CAACTGTGAG CCAAGTGGCC	2580
15	AATGTGACT GTTGTGTGAG GTGACCAATCT GGTACATTTGA GGGAACTGT TGTGCTTCTC	2640
	CACTGTGAA GCAATCAATCT TCGAAGACCG GAGAGAGAGG GTGTGTGGCT AGTGTGTGT	2700
20	CGTGTTCAC TTTCCATGCC TTTATGTGAC CCAATGTGT GTGCTGTGCA GAAAGAGAG	2760
	AAGGGGCTT AAGAGATGAA GGGTGAATAT GTATTACTTA TCCATTTCTG AATTAACATTT	2820
25	TGTATTTCTT AAAAAAATA AAAAAAATCT CAGAGGGGGG CCGGATACCC AATATGCSK	2880
	AAATGAGAG	2888
30	(2) INFORMATION FOR SEQ ID NO: 62:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1851 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 62:	
	CAGTATGATA ATTATTAATTT ATAACTAATTT CTGATTTCTT TTCAATTAAT AGTGTCTGTA	60
	GTTCCTTAAT AAGTTTGGCC AACTTCAATTT GCACTGTGTC CCAACTTTTG GACTGTGACTTA	120
45	TCCAGATCTC AGTGTGCTAT GTCAAAAAAC TGCATCAACG TTTGTGTGTA AAGATCTGTT	180
	TTGCGAATAT ATATTAATG TATCTTAATG GATGAAAGAA GTTTTAAA CAACACATTT	240
50	GTCTAGAGCT TCAATGACA TTTCTCTCTA AAGTTTCAA GTCAAGTGT TTTCTGAAAGCA	300
	AAGTGTGACA AATTGATGAG CACTTATTT AAAAACTTGA TAAAGCAAT TCAAGACTTA	360
55	CAATCTAATTT TCTTCAACCG AATTGAATTT TCCAAAGACA GTGCTTCTTT AAATGGGAGC	420
	CTGAGGGCAC TGTGCTTCTT CGTGTCAATA GACACTGCCA AACATTTAAA CCGAAGCTTA	480
	ATTGTCACTC TCCAGACTT TTTAAGCAA TCCAGACTT GTCTGTACCA GAGAAACTCA	540
60	CTCCAGAGAC AAGAAAGCAA AAAAAAATA ACTAAGATG ATGAAAGAGC AACTCCCATTT	600

5 AAAAAAGGCGC GTGTATGACG TGAATGAGGAG CACACTGTAG ACAGCTGCAT CAGTGACATG 660  
AAGACGAAA CAGGGAGGT CCGACGCCA AGCAGCCTT CTGACATGA GACGAGAC 720  
TCTCCATTA TTGATCAGG AACTGACAA GATCTCTCTT CCGCTGAAA TACTTCTGTT 780  
AAGAATACC GAATGGAAT TCCATCTTGG TTTTACAGAG ACATGTGAAA TATCAGTCA 840  
CAGCATCGAG AAGACATGC CAGCATGTT AGTATGACG ATTGTAAAGA ATTTAAAGAC 900  
CTCCACTGTT CCAAGATTC TACCTTAGCC GAGGAGAAAT CTGATTTGCC TTCTACTTCT 960  
ATCTCTCAG TTCTGTCTGA CTTAGCTGAC TTGAGAAAGT GTGATGCCA AGCTTTGCC 1020  
TCCAGAGACC CTAGGTTGC TTATCTCTC AGTTGTGCC ATTTCAGAGG ACTCTTTAGT 1080  
CATATGCCAG AACATGACAT TTATGATACC CTGTGTAGGA CCAATTGATC TACATTCAT 1140  
GTCTGACAA GATATCTGG CAAGGAAAC CAAAGCTGCTT CTTGACATTA GTGTAGCAT 1200  
GTCTACTTTT AAGTCCCTCA CCCCAGGCC CCATGCTGTT TGTATAGTT TTGCTTATTT 1260  
GTTTTGTGC TTCAATTTGT CCAATGCTCT CTGCTTGAT GCGAAGTAG ATTATAGCC 1320  
TTATTTCTTG CTCAGGCAGA ACTCCAGATG AAAAAAACTT GCATCTTCAG TATACTTCT 1380  
AAGGGCAAT CAGATATGAG ATATGTTTAA TGTATTTAAG AGTTCACTTT AGTGGCTTTC 1440  
ATTATATAG CTGTCTGG AGAGACGGG TTGCTTAGCC CTGTACATG TAAATTAAAC 1500  
TTACAGCATT TTACTGTGT ATGATATGTT GTCTCTCTTG CCAATTTTGT ACCTTATAGA 1560  
GGCAGATGC CTCGATGCC TGTGTTCTT ATTATCAAAA TTAGTTTAC TTGTTAAGG 1620  
ACACACCCA AGAATTGTA TTCTGTAAAG AATCTCTTT AGCTGTGCC TGGCAGTATA 1680  
TAAATGTCG TTTATTTAAC AGATACCTG TCGAGGAAT AAGGCACT TGAATTAATA 1740  
ATATTTTAT TATTTTAT GACATGACTG ATTGATGCT ATCTGTGCA CTTAATTAAA 1800  
CTGATTGTA TGAATTWAA AAAAAAANA AAAAAAANA AAAAAAAAA A 1851

## (2) INFORMATION FOR SEQ ID NO: 63:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

55 TCCATATGCTG ATGAGGCTCT TCGCTGGCAG GCCAGCTCTT TCGCTGCTGA TGACCTTGC 60  
ACGAAAAATG CCATCATCT GAAACATTC ATATGATATC GCGTGATCAT TGACCTCTCT 120  
60 GGCAGGCCA CAGATTCAT TATGAATGAA TATAGGATC GTATAGATCAG ACGAGCCAGC 180

5 TTCTGTGATG AGCCTTCAG AAGGACTTAA GAGAGTGAC TGAATTTGCG TAAACCCCTTT 240  
CTGTCCAGG ATGTGGAAG CTACGATCCA GTTTTGAAAC CGGTGCTGAA CCGTGAAATG 300  
CGCGAAGAC GGGGAGAGT CGTGATCACT CTGGGGAGC AGGACATAGA CCGTGTCCCA 360  
TGTGTTGCTA TCTTCTGTC CAGCCGGGAT CCAACTGTG AGTTTCCACC AGATCTCTGT 420  
TCCCGGTTA CTTTGTWAA CTTCACAGTT ACCGTTAGCA GTTTACNAG CCAATGTCTA 480  
AATGAATGAC TTAAAGCAGA AAGACTGAT GTGAGCAGA AACGATCTGA TCTTCTTAAA 540  
CTTCAGGGG AATTCAGCT CGTTTGGT CAGCTGAAA AATCTCTACT ACAAGCTGTG 600  
AACGAGTGA AAGGGGCAAT TTGTGATGAC GACACGATCA TACCACTCT GAGAGACCTG 660  
AAGAGAGG CTGCAGAGT CACGAGGAA GTTGAGGAGA CCGACATTTGT CATCGAGGAG 720  
GTGAGAGCG TGTCCAGCA GTACCTCCCG CTCTCCAGC CTTGAGCAG CATCTACTTC 780  
ACCATGAGT CCTCAGCA GATACACTTC TTCTAGCAAT ACTGCTTCCA GTTTTCTCTG 840  
GACATTTATC ACAGCTCTT ATACGAGAAC CCGAAGCTGA AGGTTGTGAC CGACACACA 900  
CAGCGCTGT CCAATTAAAC AAGGACCTC TTCAGGTGCG GTTTTAACCG AGTGGCTGCA 960  
GGCATCTGC ATCAGGACCA CATTTACCTTT GCCATGCTGC TGGCAGAT CATACTGAG 1020  
GGCAGCTGG GCGAGCCAC CTAGCATGCA GAATTCGAGC ACTTCTTTCAG AGGAATATGAG 1080  
ATTCTCTGA GTCTGCTTC CACCCCCAGG ATCCAGGGCC TCACTGTGGA CGAGGGGAG 1140  
CGCGTGTGA GCTGTAGCTG CTTTCCCGCG TTTTAAAGACT TGAATTCAAA GTTTTACGCA 1200  
GACGACAAAT TTGCAATCTG GCTGACAGC AGCTCCCGCG AGCAGACTGT GCGCTACTTC 1260  
TGAATGTAAG AACACCTGC AACACCAAT GGCAGAGCCA TCCACCCGCT GCTCTCTATC 1320  
CAGGCTTTCC GCGCGATCG CCGTTTGCCC ATGCCCCACA TGTTTGTTC ACAAACCTT 1380  
GGGAGTCTT TCAATGTCCAT CATGAGCAG CCGCTCGACC TGACCCACAT TGTGSCACA 1440  
GAGGTGAAG CCAACACTC TGTCTTAATG TGTCTGTGTC CTGTTATGA TGCATGTGA 1500  
CATCTCAGG ACCTTCGAG CAGGAGAAC ACCGAGATCA CTTCATTTGC AATCGCTCT 1560  
CGAGAGGCT TTAAACAGC AGATTAAGCA ATAAACAGCG CTGTAAAGTC GCGCAGGTG 1620  
GTGATCTGA AGAATGTGA TCTGCCCCA GGGTGGCTGA TCGAGCTGGA GAGAAGTTG 1680  
CATTTCTGC ACCCGATCG CTGCTTCCA CTCTTCTCTA CCAATGAGAT CAAACCCAG 1740  
GTGCTGTGA ATCTGTGTC TCGGGGCGC ATCTTTTGT TCGAGCCACC GCGAGGCTG 1800  
AAGGCCACA TCTGAGGAC GTTCAGCAGC ATTCGCTCT CATCGATATG CAGTCTCTCC 1860  
AACGAGGCT CCGCTTTGTA CTTCCTCTG GCTGTGTTTC ATCGATCAT CCAAGAACG 1920  
TTAGTATAG CACCTCTGG GTGTCTAAG AGTATGAT TTGAGAGTC TGACCTGCG 1980

TCATTTTCCG ATACCGTGA CAGTGGCTG GATGACGCG CAGAGCCAG CAGAGACATC	2040
TCACCGGATA AGATCCCGTG GTTCGACATA AAGACCTTAA TGGCCAGATC CATTATGCG	2100
GGGCGCGTGG AACACGATTT TGCACAGCTT GTGCTACGA CCTTCCTGGA GCGGCTGTTC	2160
ACACCCAGGA GTTTCAGAG TGAATTGAG CTGGACGCA AGGTGACGCG ACATTAAGAC	2220
ATTTCAGATC CAGATGGCAT GAGAGCGAGA GAGATTGTG CAGTGGGTGG AGTTGCTCCC	2280
CGACACCGAG AGCGCTTCCT GCGTGGGCGT GCGGCACAG CCGGAGAGAG TCCCTCTTAC	2340
CACACAGGTT GTGACATGTA TGAATTAAT GCTTGAATG CAGATTTGG AGCATGAGCA	2400
CGACCTGGCC TACCGAGAGA CTGAGAGAGA GAGAGAGCA GACTCAGAT CCGAGCGCGG	2460
CGCTCGTGG ATGCGGACAC TGCACACGAC CGCTTCGAC TGGCTGACG TCATCCCCCA	2520
CGACCTGAGC CAGCTCAGCG GCGCCGTGA GAAATCGAG GATTCCTTGT TCGAGCTTCT	2580
TGAGAGAGAA GTTGAAGATG GCGGAAAGCT GCTTGCAGAG GTTGGCGAG AGCTTGCAGA	2640
TTGTCGAGG GTTGCAGAG GAAAGAGAA GAGAGACGAC TACTTGGCA CGCTGATGTA	2700
CGAGCTAGTG AAGAGGANTC TGGTTCGAG CTGATTCGAC TACAGGTGCG CTGCGGCGAT	2760
GAGCTCATC CAGTGGGTCT CGAGCTTCAG CAGAGAGATC AAGAGCTGCG AGAGCATGTC	2820
ACTGCGAGCT GCATCTGTG GCGGCAGAGA GCTTAAAGAC ATTCAGATGT GCGTGGGTGG	2880
CGTGTTCGTG CCTAGAGCGT ACATCAGCTG CAGCAGGCGAG TATGTGGCGC AGGCGAGAG	2940
CTGGTCCCTG GAGAGCTCT GCTTGAAGT CAAAGTCAGC AGCTGAGAG GCGCGAGCTT	3000
TGACGCTGCG ACCTTCGAG TCGACGCTTT GAACTTCGA GCGGCGAGCT GCAACAGCA	3060
CGAGCTGCA CTGTCCATG CGATTCAGC GCGCTTCCG CTGAGCGAG TCGCGTGGCT	3120
CGAGCGAGCA AAGTCGAGA AAGAGCGAG TGTGTAAAC TTACCTGTCT ACCTGAAGTT	3180
CAGCGTGA GACTCATCT TCGACGTGGA CTTCGAAAT GCTTACAGAG AGCATCTGCG	3240
CAGCTTCAG GAGCGGCTG TCGAGTCTT GTGCAGAGAG TAACTTTTC TACCTGCCCG	3300
TTTCTGTAACT AGTGAAGGTT GGTATTAGC ATTATTCAT TTTTAAATA TTGTGAAGCT	3360
CTGAGCTGT GAAAGAGAG TGGTGTCT GAGCTTCAG GAGCTGAAT GGAATTCAG	3420
GCTTGGAGCT GGTGAAATTT GAGAGAGAC CAGAGAGAT TTGGAGAGCG CAGTGGCTGG	3480
GCTCTTTGA GGAATTAAA CACTAGCAT GAAAAAATA AAAAACTTA GATTCGCA	3540
GG	3542

(2) INFORMATION FOR SEQ ID NO: 64:

(1) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 883 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(1) SEQUENCE DESCRIPTION: SEQ ID NO: 64:	
AGGTGATTTT AATGATAGCT GTCAATATA GAGAGATA TCTGTTACA TTCTGTCTT	60
CTGCATGCAC TCGAGAGCG GTTACTAGCT GACAGAGAA CAAAGATCTT ATTGAAAGAA	120
GCTCTTACAG CAGCCGACG TCTCATCTTC CGATGTGAAA GATGAGCGCG CCTTGAGGTA	180
AGTACAGGCG AACAGCAGTT CGCGTTTCT CTTCGCGGCT GATCCAGAT GCGCGATGAA	240
GCGACGCGAC GTGTGTGTC TGAATCCCG GTGCTGAGA CTAAAGCCCG AGCCCGAGAT	300
CGTGAATGT GGTGCGAGCG ACTGAGAGAG GAAATTCAGT CCTTATCCG GTATGTGAG	360
AACAGAGAA ATGTGAGCA CGATGTGTC GAGTGAATC CCAAGAGGA AGGAACTCGG	420
TGTTTGAAA AATGCTGATA TATCCATGAC CTGCTGAAT ATGAGTTTGA CATCGAGTTT	480
GACATCTCTA TCGAATTC TACTACTGCC CGAGAAATG CAGTTCCTGA GCTGATGGA	540
AAGAGAGCA AGATGTACAG GGTGAGAAA ATATGCTGTA CGATCATTTT GAAACTTTG	600
TGGGCGCAGG AATGTGCCA AATTGAGCT AGCTCATCTC ATGGCTTGG GACTGGGTTCC	660
ATGGTGGCA GTGGAATCC CTGATCTGAT TCGAGAGGCG GTCATTCAGC AACAAAGAAA	720
ATGCAGACA TGAAGATTA AGCGACTGAG CAGAGGCGAG GCGACCTTGG ATTAGCTAGG	780
ATATCAATTT CCTGTGATC AGCTTAAGT CATTCAGTG TTCCCCGAGC AATCTTCAGT	840
CTAGTTGTA CTAGATATG CAGTGAAT TTGGGAGAAA ACA	883
(2) INFORMATION FOR SEQ ID NO: 65:	
(1) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 1541 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
(1) SEQUENCE DESCRIPTION: SEQ ID NO: 65:	
GCGAGAGAT GCGCTTACG CTGGGCTCAT CTGGCTACAC AAGAGATCTA AAGCTTTGCA	60
GATTCCTTGG AAGATGCGCA CCGGACAGAG CCTTCAGCA GAGAGAGAAA ATACATTTT	120
TAAAGCTTGG GCTGTAGAGA CAGGAGATTA CAGAGAGAG GTGATGAGC CTGACCGAGC	180
TAAATGAGAG GCGCAGTCC GCTGTGCTCT CAGTAAAGAG AAGAGATTTA AGCTGATGTA	240
TGATGGAGCC AAGAGAGTCC CGATGAGGCC ATGTAAAGTA TATCAAGTGT GTGACATCCC	300



5 TACAGCCGAG GCGTCGATCA TTAAACGAGG ATCCACAGGG TTCTGCTGCTT GCGATGAGAA 360  
GATATATGAT CTGATGAGAG AAGATGAGGA AGATGAGCTG GATGAGTGG AGCAACATGT 420  
TCCATCCAG GACACCTTCC CTTCTTGAA CATCATGCTT TTCTCCATGG GCGAGCCAG 480  
TGTGGGCAT TGCATGTGG GCAACTGCG CCCGAGGGA GTGTGGCCA AACTGAAAC 540  
CTTGGAGTG GAGTATGCC AGGAGCTAT AGAGCCCTTC TTATGCTCTC CAGAACTGTG 600  
GATCAGCTCT CTCCCATGA CTGACCTGGA CATCAGTTT CAGTACCGTG GGAAGAGTA 660  
CGGCGAGAC ATGACCTGA GCAACCTCA GCGCTGCCGA CTCTTCTATG GCGACTGGG 720  
TCCATGCTT GACGAGAGG AGCTCTTTGG TCCGCTCAGN CTGGAGGAGG TCANATTCCT 780  
AGTCTCTGAG CATATTACCA ATGAGAGCA GAGCTGTTC ACTAGCAGCC TCGTGGAGT 840  
CATGAGCAG GCACTGATCC TGGAGTCCAG CGGTGATGCC ATTATGCCA TCGAGCTGTG 900  
CCAGTGCAG GTGTACTGCT CTGGGCCATG TCGCCCATCA CTGTGTTGTC CCAACTGAT 960  
TCAGAGACA AAGAGGTCA AGCTATTTTG TGTGGAGCA TTGCTTAGGG ATCTCATTC 1020  
CCACGAGAA GGACAGATAG AAGAGGCC ACCGTTTGA ATCTACTTAT GCTTTGGGA 1080  
AGAATGGCA GATGGGAAC CATGGGAAG GAAACTATC TTGCTTCAGG TCATTCGAT 1140  
AGTGGCTGG ATGATCTAG AGATGTTTC TGTGTATTC ACAGCTGCT TTGATAGTG 1200  
CAGTGTCCC CTGAGATCT CACCCGAGA CATCAAGAT AACTGCTTG CTGAGCTGA 1260  
GCACTGTAC GCAATCTTC AAACCGAGA GAGCTGGAG CCACTGAGC CCACCCGAG 1320  
CATGCAACTG CCGCTGCCC TCGCTGCCA GTATTTGGA ATGCACTCTT CTCTCTCTC 1380  
TTTTTTTAA TATCTACAT ATGATTTT TTATTTGTTA GATTTAACA GCTTTTAAAT 1440  
CTCTGTTTT TGTGACAGT TTGAAGTTT GTGATTTCC AAATATGCTT AGATTTAAG 1500  
CTGATTTAAT TTATGGAAA AAAAAAAAAA AAAAAAAAAA A 1541

## (2) INFORMATION FOR SEQ ID NO: 66:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 712 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

55 AAAAAAAAAA TGTTHAAGG TCGCTGCCA GCGCGAGAG AGTGTTTGCT CCGCTGGAG 60  
AAGGCTCTGC TCGCCCTTGA GATGCTCTC CTGCCCCACC GATACTGGCA CTTTAAAG 120

5 GAGCTGACC GCACTGTCT CAGAGGATTT GCGCCCGAGA AGATGGGGAG TTCTGTCTCTG 180  
CCCTTCTGT TTCTGCTGAC CTCACCGAG CTAGAGGGA GGTGCATTTCA GGTAGATTTT 240  
GCTCTCATT CAAATTTCTG GCGCTTTGGG CGGAANACAG CCACTTTTGG GCGTGTGGG 300  
GAGACTCTTC CAGACCGGA ACCCGAGAG GAGACAGAG CTCGCCATC CTCGACGCC 360  
AGGCTCTGG CCAAGCTGAT TGGACTGAGA ATTGCGCAC AACCAATTTG ATGCTGGCTG 420  
GAACGAGAG CGAANAGCC TGGCTTTCTT CCAATGTGG AGCCCTTCTC TCAGCCCTCT 480  
TGTCTCTTG ACTGAGTGA ATTCGCCCA GGTGCCACA GCTCTTGAC TTCAATTTCT 540  
ATATATTCAG AGATTTGGAG AGTATATCAG AGATATTTT GGAAGGAGT TGGTCTATGC 600  
AATGTAGTT TGGATCTTC TTGAAGTTT AATGTTTTTA TTAGAGATTT TAAGGAAT 660  
AAGGTCTAC ATATCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 720  
AAAAAAAAA AA 732

## (2) INFORMATION FOR SEQ ID NO: 67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 629 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

35 TTAAAGATTT CGGCGCATC CGGCGAGTA ACATGACTAA AAGAGAGCG GAGATCTGG 60  
GCTGTGCTCT AGAGATCGAT GCGTAGAGG AGAAGCTGTC CCAATGTTCG AGAGACTGG 120  
AGGCTGTGAA CTCGAGACT CACAGCCGGG AGCTGAGGCC AGAGGCCAGG AGGTCCCTGG 180  
AGAGCGGAA AACAGCGTA ATGAGCAAG CTTCCAGTA CAGAGAGAA CTGAAGTTTC 240  
TTGCGAAGA GAACCGGAG AACATGCTGC TCTGTGTGGC CATCTTTATC CTCCTGAGCC 300  
TGTCTATGC CTACTGACC ATGTGAGCTT GGCATTTCC CACAAACAG ACAGGCTTCC 360  
ACTTGGCCC TTGCTCAGGA TCAGCAGGC ACTTCAGGCC TCATAGGAC CAGAGTGTG 420  
GCGTGTTCCT CTCGACACT ATGTTTCAG CATGCTTCC TGGCGGCCA GCGCTTGCCT 480  
CCCTGGCTG CTGGGGGTT CCGGCTCTCC AGAAGGACAT GGTGCTGCTC CTCCTCTTAC 540  
CCCAAGGAG AGGCAATTA GACACAGAG CTGAANAAA AAAAAAAAAA AACTGTGAG 600  
GGGGGCGCTT ACCCATGCC CCTTCTGCTG 629

60

## (2) INFORMATION FOR SEQ ID NO: 68:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1751 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

10 CTGCTACCCG GCGCGCCGAG GCTGCCAGAC GCGTGAAGCC GAGAGCCAGG CCAAAACCTT 60  
GGTAACCCCG GCGTGCAGCCG CTGAGTCTGC GGCCTAGCGG GCGTCCCGCC GCGTCTCGCG  
15 AGCGTGAATC ATGCGAGGCC CCGGCTCGGG CAGAGGCCAC GTGTCTGCGC GCATCACTAC 180  
AGACTTCGAG CTGAGGACCC TCTCAGCCGG GAGACTCTCT CCGGAGAACG TGCCTCGCGGG 240  
GAGAGAAATC GCGGTGTGAG CGAAGCGTTT CATTCAGCAA GGGAAACTCA TCCCGAGATG 300  
20 TGTCAATGAT CCGCTGCGCC TTCAATGACT GAAAAATCTC ACCGAGTATA GCTGACCTTT 360  
GAGTGGTTTT CGAAGACAC TTCCAGACGC AGAAGCCCTA GATAGACCTT ATCGAGTGA 420  
25 GAGAGTATC AACCTGAATG TCGCCCTTGA GTCATTTAAA CAAACCCCTA CTGCTCGCTG 480  
GATCAATCCC GCGATGCGCC GAGTCTATTA CATGGAATTC AACCTCCCA AAACCTCGCG 540  
CATGATGAC CTGACTCGCG AACCTCTCAT TCGACGTGAG GATGATTAAC GAGAGAGCGT 600  
30 TATCAAGAGA CTAAAGCTT ATGAGAACCA AACAAAGCCA GTCCGTGAAT ATTACAGAAA 660  
AAAAAGGCTG CTGGAACATC TCTCCGAGAC AGAAGACAC AAGATTTGCG CCTATGATATA 720  
35 TCGTTCTCTA GAACTTAAG TTCCAGAAAG AAGCCGAAA GCTTCAGTTA CTCCAGTAGG 780  
AGAAATGCTG GTACCTATTA ATGATGAAT GGCAAAACCT CCTATCTCTT GCATTTTGA 840  
GCTGCTTTTC CTAGACTTTC TACATGATAT GAACTCTTTC AAAATATATC TACTTTTAT 900  
40 TCTACTGATC TTATTTTGA TACTAAGGAT GTCCGAATG ATCGGATAC TAAAGATCAT 960  
GCTTGAATC CATCAAGCTT GTTGATGCA GTTATCTCA AAACATCGAG CGATGCTGTA 1020  
45 AACCTTAAA CATCTGTGAG AGCAAAATTA AAGAGACATC TCGATGAATC CTACTCTTTT 1080  
GTTCCAGTTA TAAAGCTTTC ATAAATTTTC GATATTTTTC TCGAAAGTTT AAAAAAGTTT 1140  
50 AAGATGATTT TGGAGAAATC ACGTAACTAG AAATTTGTC ATGATTTGAT GCGAAAAAG 1200  
AAGTTTCAG CATCTGGA CAGGTGAGA GACTATATTA AATTCAGAAA AGAAGACAC 1260  
TGAATTTACA GATTATTTCT GAGACAGAAA TTCACTGCTG CTTTACACT AAGAAATGTA 1320  
1380  
55 TATGTTAAC ATTAATGCTG TATTTATTTT GTCTTTAGC ATACTTCAG TTTCATCAA 1380  
ATTTCATTT TCGTATTAAG ATGTATCAT TACGATTAAG AAAAATTTA CTTTAAAG 1440  
ACTGTCTCC TTGAAAAATA CCGTGTACT GAGGTGAG ATTTGTGCTA AAAATTTACA 1500

TAAATCTTT TAAAGACAC AAATGATC GATTTTCA GAAATTTA TTAAGCTA 1560

TGTTCATCT TATGCTAGG GTTAAAGAT GTGTTTAG ATCTACTGC TGTTTTCT 1620

TTTGTATCA AATGTGAT CTGCTGTAT AAATTAAGG TTAATTAAC ATCTCCCC 1680

CGAATAAAA AAAAAAAGG TCGAGGGGG GCGCGTAC CAACTTCCG 1740

NAAATGAGG T 1751

## (2) INFORMATION FOR SEQ ID NO: 69:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 508 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

10 GCGAGAAAT TATGATTTA AATGTTTG AATTTGAAA TATTGAATA TTGTACTAT 60  
25 TTGACCCAG TCAAAATCTC CATGAGAAA TACTGTGTA TTCCAGACT ATTTGTAAA 120  
ATTATCAAT GCAATATCAC AGCTGTGCA GACTGTATA CGAGTTGCGC AATCAAGCCA 180  
30 CAGTAAAA TTGAAACAAA AAGATCGAC AACAAAAGC CCAAGGCTG GCGTCAAGAA 240  
GCTGTAGCT GTACTTACT GTAAATCTC ATGCAGACT GCGAGGTGTA GTGTGCAAT 300  
35 GAGAGATG AAGCTTAGG CAGTTTAG CTCTGAAGC ATTCAGAAC ATTAAGACAA 360  
CGAGGAAAA CTGAGAGAG CCGAGCTT GAGCTGCT TTACATGAG CTCTGTGTA 420  
GAGAGATG AAGCTTAGG CAGTTTAG CTCTGAAGC ATTCAGAAC ATTAAGACAA 480  
40 ATCGGCAA GCAATTAAGA AATCTGTA CTCTAAGTG TTCTGACCC AGAAGAACT 508  
CTACTCAC TCGCTTTAA AATTAATA

45 (2) INFORMATION FOR SEQ ID NO: 70:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 245 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

55 TACTTAGAC AAAGAAAT TTCCAGAT TGTAAATTC TGAAGAGA ATTTCTGA 60  
GATTCAGAT TTGCTGTGT CCGAGGTGA TGATGAGGC TGTTTTCCC TGTGTCTTT 120  
60 TCTCACTACT CAGCTCTCT CTCTCAACT GTCTGTTGG CATATGATG TGTCTACTG 180

GCATTTCCTT CACTGATACA AGGAAACTG CAGGTTTAA AAAAAAAAAA AAAAAAAAAA

240

NENCG

245

5

(2) INFORMATION FOR SEQ ID NO: 71:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 361 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATGTTTCCTCA TGAGATGCA CTGTGCTTC TCGAGTATT OCTGACGTT CATAGTACT

60

CCACACGCA CCAGCAATAC AGTACTAC CTGTGGCTT GGATCTCAGC CAGATGGCT

120

GGGAGAGGA GAGCTGGGC ATGTACCTTA AATCTGTTA CCAAGGAAG ACTCCAGAG

180

25

TGAGACAG TAGGACTTC CTGAGAGGT GGTACATGG CTCTGTGAT CATACTTTT

240

TTTTTTTTT TTTTGAGATA GAGTTACG CTGTGTGCC TGGCTGGAGT GCAATGGTC

300

GATCTGACT CACTGCAAC TCTGTGCTC CCGGTTCAA GTGATCTCC TGGCTGAGC

360

30

T

361

(2) INFORMATION FOR SEQ ID NO: 72:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 713 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

AGGATCAC ATAGAGAAC ACTGTATGA CATTTCGGT TCGTCAGG ACCGAGACA

60

TTGATGAGT TTTGTGTGC GTTATATAT TTTCTGTTA AAGTTTGTT

120

TTAGTAGAT GATATTTTA TAGCTGGAG TCGTTTGGAA CTATTAAGAT GTCACTACT

180

AACACATA CCTATGTTT TGTTTTGTT TGTTTTACG TCAATATAA TCAGGAGAG

240

TTAGCGAGC ATTAGCAAT TAGATCTTC TTTTATTGG TCTCTAAG ATATGATGT

300

TCCCATACA CCAACAAAC AGCAACAAA ACATTCATA AATACACTT GATAGACTGT

360

AAGCACTGC TTACTTTGT GTCCAAATA TTTAGTGTG ATATATATAT ATATATATAC

420

ACACACAC ACATATATAT TCAACAATA AAGCAATAA TACATGCAT TTCACATTT

480

60

GTCTTTCCTT GTTACGATT TATAGCAGA ACTGTATGAC AGCTTTAGT GATCTAGCA

540

TATGTTAAT TCAATTAAT GTAAACAGA TTAACACAA CAAGAAACT GTCTATTGA

600

5 GTGAGTCAT GCTTCTTAT ATATAACTT GGTTCGGTT ATCCATCAA TCACACTTA

660

TACTGTATC TGATCTTTA TATTAAGAA TACTGTACTT ATAAAAAAA AAA

713

10

(2) INFORMATION FOR SEQ ID NO: 73:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 862 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

GAAAGTCAGA GCTGTCCAT CCTCAGAC CTTTAGATT TCGTCCAAAT TAGAAAGTCG

60

GGGACTATG GTTCTGGCA ATCAGAGTC TGGAAATGG CTCGAGGC TCTTGATAGT

120

25 GAGACAGTGG TCATCTTACC AGACATGCAT CTGATTTTAA GCTCAGGCT AATCCACAT

180

CCTCGGCCCT GCTATGATT AAGAAACAAA ACCAAATCT GCTTTTAGT TTTAGGAAAC

240

30 CTGGATAGA CAGTATTTT CAGCATCTT GGATAAAGCA GTTCTGCAT TTTAAATGG

300

GACTGCAGAA GTGACTGTCT ATAGTTTGA AATACAAAA ATGATATGTT TGATCAGAAA

360

35 AAGAGCCCG TCGCTGGAC TTGGAAAGT ACTGAGGTC ATACCTCTTA TGAGAAATG

420

TAGGCTCTGT GATGTTAAC TACAATCAG GTTAGGAAG CATATGACAC CTTTGTGCA

480

ACTAAGCTTC ACTAGGAGA CCTGTCTCA TAGAAGATA TCGTTTAAA GTATCAATTT

540

40 TCCACAGTCG ATGATGAGA AAGTTCAAT TCACACGAA TCGTATAGT CACATGAC

600

AGCTGACAT ATATACANT ACAGTTTCT GTAAACAGAA GTTCTTCTC TTCAATTCA

660

CGAGTCAGC AGACATAAA TATTGCATGT TTCACTTAG AAACGATTC ATTTAGAAA

720

45 CGAGATCTGG ATATTTTTC AGGGTAGAAA TGAAGGCTAT TTCGGCATT CTTGCTCAA

780

AATGCAATAT ATGTACHTA AGTATAAAA AGGCTCTCT TCACTCTTT TGTTCGTAG

840

50 CATTGGCTAC ATACTCTG CC

862

(2) INFORMATION FOR SEQ ID NO: 74:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4602 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: Linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

5	GGAGGGGGGCG GKGGGAGACA GGGGGGAGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG	60
	GGGGGGGGGG GAGGAGAG GGGGGGGGCG GGGGGGGGCG GAGGGGGGCG GAGGGGGGCG	120
	GAGGAGAGAG GAGGAGAGAG GAGGAGGGGCG GAGGAGGGGCG GGGGAGAGCG GGGGAGAGCG	180
10	GGGGGGGGAG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG	240
	AGGTCGAGGAG CGAGGGGGGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG	300
15	AGGGGGGAGT TGGGGGGGCG GAGGAGAGAG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG	360
	GAGGAGAGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG	420
20	CTATCGAGAG AAAAGAGTAC AACTGATGAA GATTCGAGAG AGAGAGAGAG GAGAGAGAGAG	480
	GATTCGAGCG TGGGGGGGCG AGGTCGAGT GCTGAGAGTCA GAGTCGAGCG GGGGGGGGCG	540
	AGTATGAGAG GATTCGAGT TGGGGGGGCG GATTCGAGTCA AGAGGGAGAG GTCGAGGCGCG	600
25	AGAGGGGGTTC GATTCGAGCT GATTCGAGAG TCGAGAGACA ATTCGAGAGCA ACTTCGAGTAT	660
	GATTCGAGAG AGAGGGAGCA GTCGAGAGCT GGGGATGATCA GTGAGAGAGCA AGAGGGAGAG	720
30	AGAGAGAGAG TGTTCGAGAG TCGAGAGAGAG ATTCGAGTCA AGAGTATTC AGAGTATTC	780
	AGTCGAGAGCG GGGGGGGGCG AGGGGGAGCG GGGGGGGGCG GGGGGGGGCG GGGGGGGGCG	840
	AGAGAGAGAG AGAGAGAGAG GAGATTCGAG GTGAGAGAGAG AGGTCGAGCT GAGAGAGAG	900
35	GAGATTCGAG TCGAGAGAGAG TCGAGAGAGCT GCGAGAGAGAG GAGAGAGAGAG AGAGAGAG	960
	GAGAGAGAGT GATTCGAGTCA GAGAGAGAGAG AGAGAGAGTCA TGTTCGAGTTC	1020
40	GAGAGAGAGAG GATTCGAGTCA TGTTCGAGTTC GATTCGAGTCA GATTCGAGTCA GATTCGAGTCA	1080
	TGTTCGAGTTC TGTTCGAGTCA GAGATTCGAG TGTTCGAGTTC GATTCGAGTTC AGGAGTTC	1140
	AGGTCGAGAG AGGAGTTCCT GCGAGAGAGCG AGAGAGTCA GAGATTCGAG GAGATTCGAG	1200
45	TGTTCGAGTTC GATTCGAGTTC GATTCGAGTTC TGTTCGAGTTC TGTTCGAGTTC GATTCGAGTTC	1260
	GAGAGAGAGAG AGAGAGTCA GATTCGAGTTC TGTTCGAGTTC TGTTCGAGTTC GATTCGAGTTC	1320
50	CTTCGAGTTC GCGAGAGAGAG AGAGAGTTC GATTCGAGTTC GATTCGAGTTC TGTTCGAGTTC	1380
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG GATTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG	1440
	CTTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC	1500
55	ATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC	1560
	TGTTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC	1620
60	GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC	1680

5	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG GATTCGAGTTC GATTCGAGTTC TGTTCGAGTTC	1740
	TGTTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC	1800
	GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC GATTCGAGTTC	1860
10	AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	1920
	AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	1980
	AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2040
15	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2100
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2160
20	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2220
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2280
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2340
25	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2400
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2460
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2520
30	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2580
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2640
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2700
35	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2760
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2820
40	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2880
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	2940
45	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3000
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3060
50	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3120
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3180
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3240
55	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3300
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3360
	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3420
60	TGTTCGAGTTC AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG AGAGAGTTCAG	3480

TTTCTTCAC ACTTATCCCT TGTCTCGTA AATTCTATT GAGTGGTTA GTCAACAGT 3540  
ATTATTAGCA CCTACAGAA AGCAACTGCG ATTTTAAAA ATCTTTCTGA GATCGGAGA 3600  
5 AATGTATCTT CCTTTCTTAT ACCTCTCTCC CACCAAAAA ACACTAGATT AGTTCTACTA 3660  
ATTAGAAACT TCGTGTACTT TTCTTTTCT TTAGGGGTC AAGGACCTC TTATTAGCTA 3720  
3780 CCAATTGCGT ACATTAAMTT ATTGCAGAG TTTCGATAC TAAATATTT TTATTAGACT  
3840 TTATATTTTT CCTTTTGATA AAGGATGCT GCATAGTAGA GTGGGTGAA TTAAACTATC  
3900 TCAGCCGTTT CCTCTCTTC CTTCTCTC CATATGCTC ATGTCTCTC CAGGAGCTC  
15 TTTTAATCTT AAGTCTTAC ATTCTAGCT CTATGTCAA TTCTGTACC TTTTAAATA 3960  
CTCTTCCAC TCATATATTC CATCTCGAT TCGTGTCTT AATTTCTGA ACTGTAGTG 4020  
4080 AGATACAGCT ATTTAATATT TCTGGAGAT GTCAATCCCT CTCTTTGTG GTTCCCGAG  
20 GTTGTATTGC GTACTGAGA CTCCTGATA TCGTTCAGAG ATTTAGGCA AACACTGGC 4140  
ATGCCCTGG GAGTACTGG AGTAATAAA AATATCGAG GTATAGATA GCATCCAGT 4200  
25 AGACCACTG AACCTCTTT GTACCTGTT GGGGAAAAG TATATAGT GTACTAGCA 4260  
4320 TCTAATAG ATTATATAG TCTGTGTTT TGAATAGCA TTTTCTCTC CTTTGTGTT  
4380 TTTCCCACTT TCCATGTA TCAGAAAT TGAACAATG TATGTATCA ATTTAAATA  
30 TTTTATTTCT TAAAGCCTT TTTTCCCTT TGTATGTGC AAGACCTTC TCTTTTCAG 4440  
4500 GGAGAGAG GTAGTACTT GATATAGT TGAAGAAGT ATGTAAAGG AATTTATAT  
35 AAAAGGATA CTTTGTCTT CAATCTTTG TTTTCTCTTA TTCTAGTAA GCATATATA 4560  
4602 AATAAATAT GTAAGAGA AAATAAAG TTGTCTCAT G

## (2) INFORMATION FOR SEQ ID NO: 75:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1255 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CGCCGCCGG GCGCGCGGT TTCTTACA AATAAACA ACCCGCACTG CCCAGCGAG 60  
CGTTGCCAT TTCAAGTGG TCCCTGGG GAGCTAGCC TCACTGTAT GATCTGCA 120  
55 AGCGCACTT TTTATTTTA TTTATTTT ATTTTTTT TAGCATCTT TTGGGGCTC 180  
ACTCTCAGAG CAGTTTTHA AGGAGACA GAGCCGAGC CTGCTGAT TCTATGCTT 240  
60 GGTGTGTTAT ATAGAGTAA TTGCTAACT TGAATTTTA TCTCTTAAAC CAACTGTG 300

GCATAAGAT ATTGACCGT TTCCAAAAT CAGATCTGC CTCGCGGAT AATATTTCG 360  
CAGGATGAG TACTCTCTG CACCACTCG AAGTCCAGA CAGAGGTTT TACACATTC 420  
5 TTAGCACTGA ACTCTCTGT GATTAGGAT GATCTGTCC CCTCTCGAT GAACTCTC 480  
TGATGATA GGCCTCCAGC AGCTACTTT GAGGGGAACA ATCAGATGCA AAGCTCTTG 540  
10 GGTGTTTTT TAAATACTA GTGTACTTT CTGAGTAGCC GCGCTTCAAC AGCGTAGTC 600  
CAGCCCTGG TCGTTGTAG AGCGAGTGC TTGCTCAGAG CCAATTTTC ATTTCATCA 660  
720 TTACTGCTT CACTGCAATA GTCACTCTT TGAATCTGG GAACCAAAAT GTGATGATA  
780 TATAGACTTT ACTATAGCC ACAGTTCTC CCACACCTA GTCTTCGAA TGTATATAT  
840 TGAATATCT AGAAATGCA TTCTATCAAT TACAGATTC AATATITGCA AAGGATGTC  
900 TGTCTTTCT CCGAGCTGC CCGTTCCCT TCAATGAAA ACCACACGG TGCATCTCT  
20 TGTGATGCA GCGCTATGCA CCGCAGGCA CCGTGTATG CACTCCCGC TTGTCTTAC 960  
ACAGCTGG GCGTGTAG CAGTCTGCT TTTTCACTT AATAATACAG CTTCAGAGA 1020  
25 TTTTGTATC ACATTAATA TCCACTGCG TCTTTTGTAT GGCACATAA TAACTACTGC 1080  
ATATATGCA TAGCGCTAT TTGATTAAC TAGTTCCTA ATGATGACT TTTAGTGTG 1140  
30 TTTCTTTT TTTCTTTTT GCTAGTCAA AGATGCTAT AATAATGTC CTTATCAAA 1200  
1255 AAAAAAAAA AAAAAAAAA AAAAAAAAA NCGGGGGGG CCGCGGAGC NCAT

## (2) INFORMATION FOR SEQ ID NO: 76:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 475 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

GGCAGAGAG AATGTTTGA TTCTTTTC TATTTAAGG GATCTTCTT CTGTGTGATG 60  
120 TTGAAACTT ACCTTAGTGA AGATGTTT CAGATGCTG TTGTCTTTA CTGCATAT  
180 CAGACTATG CATCTATCA AAGTGATGAT CTGTGGGATA GTTTTATGCA GGTCAACAC  
240 CAACTAG ATGTAAAGAG AATGATGAA ACCTGGACC TCGAAGAGG ATTTCCTTA  
300 GTGACTTTC AAAGAAAGG AAGGAACTT TTTATCAAC AAGAGATTT CTTTTAAAT  
360 ATGAAGCTTG AATTCAGCC TTGAGTACA AGGTACTGCG CCTCTTCTT TTCATGCCAT  
420 CTCCTTTGCA CTCACAGTG GAATATTTT GAGTCTTTT AATATCATAA GTTCTGTGA

AACTTAAGA GATTAACCTT TCTTAAGAT ACTTAACCTT GCTTCAAAAT TAAAA 475

5 (2) INFORMATION FOR SEQ ID NO: 77:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 465 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

15 TTCTCTGCG TCTTGACATG CACCCACATC GCGCGTGACC CTGACTCCCC CTAGTCAACT 60  
CAGCGGTGCT GCGATGCGGCT GCGCGCGCGCG GAAACCTGCG TTGCGGACCTCG GCGCGTCTTG 120  
GCTCTGCGGCT TCGTGGGCTT GCGCTCTGTC GTGCGCGCGCG CCGCGGCGCG GCGTCTGACG 180  
TGTCTCTCGG CCGTGTGAAA CATTCAGTAC GTGCAACCCG AACACAACTT GACCGTGTGCG 240  
AGCGTCTCGG AAGATGCGCG CTTCGCGGAC AGCTGCGCCA AAGAGGCGCG GCGATGCGCTG 300  
GTGCGCGCTG CCGTGGCGCG CCGCGCGAGAT CTGCAAGGCT KCGCGCGCGA CACCGCGCTTC 360  
TTGCTGCGCG AACCGCGCGG CCGAGCGCGC GCGCGCTGCG TTGCGCGTGT GGTCTGTGCGG 420  
30 GCTGCAACTT TCAAGAGAA AATGCTGTGT GCGCGCGACG AAGAA 465

35 (2) INFORMATION FOR SEQ ID NO: 78:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1907 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

45 AATGCGAGCG CACTTACAGA TTCTTAATGA ATTCTGACG GTTGCAGAA GAAATTAAGG 60  
AAGACAACTG GAAACAGATC AAGAGAGCT AAGTGTCTTG GAGAGAGATA TTAAAGAGAT 120  
GGAAGAAATG AATGACTTAT ACTCTGCTGT CAGTGAGCAT AACACAGTGC CTCAATTGCA 180  
AGCTCTCTCT CAGTACAGA GTAGTATTAT TGAATTCACA GAATACAGCG AACCTGACAG 240  
TTTCAGTGGC AGTCTCAGA CAAGAAACA GCGTGTGTAT AATAGCAAGT TTAGCATTAG 300  
ACGAAGACA CTTACTGCTC AATTGAAAGA CTTCAGACAG TGTACTCTTT CTACAGAGAT 360  
GCTCTGTATC TGAATAGACA GTGAACTGC AAGCAAGTTG GATGAAATTC AGGAATGCTT 420  
60 GTCCAGGTTT ACTGATATA ATTCAATAGC AACTTTAGCC AATTTGAT ATGCTAGTGA 480

TTCTTAATAT GATTCAGATA TACCTCTAGC TATTGAATTT GACCGGATTT GTACATATTT 540  
TGCATTTCT GAGATTACAA AAGATTTAAA AGCTTAATGA TATGACACTG TCAATCAGCA 600  
TGCATGTGAT ATTCAATTAC CTGAGATAGA AATGACATGC AATTGAAAA TCAACTGTAT 660  
5 CAGTGGAGAT AATTAACATA AAGACCTGTT AACTGACATG GATTATGAGG GCACTGTAT 720

10 (2) INFORMATION FOR SEQ ID NO: 79:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

15 TAAATGTGCG TGTATTAAAT TCAAGCGCTC TTGCAAGTAC GATTGCGCTT TCGCGTGC 960  
AATGCACTGT GTCACTACT ATGATCTTGC TAACTATAAA CAGCAATCA TGTATTGCAA 1020  
AAGACAGCTT AAGACAGCTT CTATACAAA GTTGTGAGT GTGAGAGAAA TTGTCTCTGC 1080  
20 CTCAAGACAG AGTCACTTAA AACTGTGAAA TGTAGAGAAA CCAATGCTGC TACGTCTCTT 1140  
CAGCGGTGAT ATTCAATGAA AAAACTTGTG AAGCGTGCCT TCGAATGAGG ATTATATAC 1200  
25 TTGTGAAAT GAAATTAATC CTCTTACT GTACTTAAA GACTTTCTTA AACTTTGCT 1260  
AACTTAAAG TTGATTAAGC TCAAAAGTGT TCTGACAAA GACGAAAG AGATGATAC 1320  
AAATGAATTT GTTATGCTG TGTCTGACG GGCACATACA GATGCGAGCT CCAATGTGCT 1380  
30 GATTCCTGCT AACCTGACG GTACATTTAA GGTCTGAGAA TTGTATGATA GCGTTAACTC 1440  
AATGCAATTT GTACTGATC CTGCTGAAT AACCTGACG CTGACATCA GAGAGAAAC 1500  
35 AAGAAATGTC AATGATATGC TCTCCAAA GTCAATGATG GTTGTGATTT TGTATTGAAT 1560  
AATTTTCTCT TTTTCTCTTT TCCCTCTTT ATGACTCTTG GACATCTGCG AATACCGAC 1620  
CACTCTTCA CCAATCAAT AACTGCAAGC AAGATGCTGC TGTGTGTGAT GTTATCTAAT 1680  
40 TTTTGTGATA GCGAAACAAA TTCTTTGATA TAAATATAAA TAAACAAACA ATAAAGTTT 1740  
AATGACCCAG AATGAGCTT GGAAGTTT TGTCAATGC NCGAAGAGAT AACTCTTTT 1800  
45 AATGATGAC AATGATGAC TAAATGTGTA CAGTGAATTA TTGTAAATTT TGTATTCTCC 1860  
CACTCTTTT GCGAATTTAA CTATCAATA TAAACAAAT ATAAAGT 1907

50 (2) INFORMATION FOR SEQ ID NO: 79:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60 (14) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

5 OCTGGGGTGT CCCCCKSGCC ACCATGTCTA TCCCTTACTT GATGAGGCAC ACTCGGATGA 60  
CCCATGACTG ATCGTTTATTA ATTGTCTCAA GGCNAAGCAC CATTATCTC CCCAAACCTT 120  
AACTTCATGG GCCAGTGTCT AGAGTTCGAG GAAGACCTTA ACAGCGGTGT GACACCGAGA 180  
ATCCTTACAC CAAGCTGAT GGGCGTGGAG ACCGTTGTGT GACATGCTC TCGATCGAAA 240  
10 GGAATGCTGC TCTCCATTAG GAGACATGA GGAAGGAGA TGGATTTCTG TTTTITTTCT 300  
TTCTTTTITTT TTTTGTAGTT GGGCTAAGT TTTGTAGTGG AACCAACTT GTTTAAACAC 360  
TTTATTTTAA ACAGTGTAA GAAGACTATA ACTTTGTATG CCATGTAGAT TCACTTCCA 420  
15 CAACCTGACA AATTAGGAG GTTAAGGAG TAATTTTTTT AAGCCACAA TAAATATATA 480  
ATACAACTGG TTTCTGCCCC TTTCTCTTTT AAGCTATTTG TAGAGTTTAT GACTAANTAG 540  
TCTGTGAGG TTTCTAGACC GAAGTACTA CACACTTTAA ACCAATTAAA AAGAACAAA 600  
20 ACTAATATGA AAGACATTTG AATCACCAG CCGTGGGATC AAGCTGGGCT GTCCACACAG 660  
AAGACAAAA CCCAACAAA CCCAGCCCTG TTGCTCTAC TGGTGCAGAG AAGAGTACG 720  
25 GGCAGCTTAA GTGGTCTTAG TATCTTTGAG GCATTTCTTA AGGAGAAAA GGAATCCTTT 780  
GATTTGTGT GTTTCAGCT CTGGATTTT TTTTITTTT CTCTCTGGG TTTAGAGAT 840  
30 TTTTITTTGAA ATAGTGAAGA ACTGAGCATT ATATGCTTTC ACTGGCTTCT TGTGCATATA 900  
TATGATGTTT TANGTGTGCA AACAGTTTAT AGCTGGCAGC TGAATGATAG ACAATAGTG 960  
35 CAATTTTCCC AGCTTGAGGA TAGAAGGAAA TTCAACATATA TATCAATATC TTTCTTCCC 1020  
ACCTTTTCCC TTTTITTTT TTTTITTTCTA TTTGATTTCT GTTACAGTGC CATAAACCTT 1080  
GTTACATATG TATATCAGAA TGTAAAGAAA AAAAATTTAT TTAATAATAT TTTTGCAAA 1140  
40 AAAAAAANA AAAAATCTGA GGGGGGCC 1168

45 (2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1285 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

55 AAAAAATCAC ATCTAAACAA AGAATCTGT CTAAGACAGT ACATCTCTG TTGAATCTGC 60  
ATCTTTCCAC AGCACTTTCT GTTTTATGGG ATGAGACTAT TCTCTCTTC ATCAAGGAAA 120  
GAGAAATGTT CAGGGTTTATA GGGATGGCAC ACTTATTAGT TCTGCTCTC TGAAGGTTTC 180

CTCCAGGACA GTTTGGTCAG AGCTGCAATT CTATGTCCAT GGTCTAATCC TTGAGTATCT 240  
CTTTTTCGCC TTTCTGTCT CAGGAATACG CTGAGAAATC ATTCGATTTT CATGCCCTTA 300  
5 GCCCTTACT GTGATTTGTT GTTTCACCTT TCATTTGCTT TAGTTCTAGA ATCACTGTTT 360  
GACTCTCAG ACTTCACCTA ACTTTGAAA CTCTCTTTTG GAGGCTTCTC ATTTCCOCCCT 420  
AATCTGTGCG TCCCTGAGCC CTAGAAATTT CCCACCAAGC AATTAATCCA GGTAGATCCT 480  
10 AAGTTGCTGG ATCTAGTTGA TATTTAACA ATATCTAGTT GATATTTCTC ATTCAGTTGG 540  
ATCCAGAAC CAGTATCTCT NAANAACAAC CTCTCATACC TTGTGACCTT AATTTTGTGT 600  
15 GCGTGTGTGT GTGGGGCAT ATGTATATAG ACAGGCACAT CTTTTTTACT TTTGTAAAG 660  
CTTATGCTTC TTTGTATCT ATATCTGTGA AAGTTTATAT GATCTGCCAT AATGCTTGG 720  
780 GGACTTTGTT CTCTGTGTA AATGTAATA CAGAAACAC CTATATATAG AGTCAATCTA  
840 GTTGTTTTAA TTGCATGA AGAAATTTT CAGATACAAA CACTAACAAA CTCTCCCTTG  
900 ACTAGGGGGA CAAGAAAGAG CAAGACTGC CATAAAAAC AATTACTCGG TGAAGATTC  
25 CATTAACAGA ATTAGTAGT ATATGAGA CAGCATCATT AAACAGTAT GTTGTCTCC 960  
TTCCAAAAA CATGTACTGA CTTCOCCCTG AGTAATGCGA AGTTGTTTTT TTTATATATA 1020  
1080 AACTTGCCTT TCATTACATG TTTCAAGTG GTGTGCTGGG CCCAAATATT GAATATGCG  
1140 AACTGACTGA TAAAGTCTA CAATAAGCA GTGTGCTTAA CAGCAACAC AGTAATGTTG  
1200 ACATCTTAA TTCAAAATG CTAAATTTAT TATAATTTGT TTTGCTAAAA TACACTTTGA  
1260 AACTATTTTT CTGATTTCCA AGAGCTGAGA TCTTAGATTT TATGTAGTAT TAGTGAAAA 1285  
AATACGAAA TANTAACAT TGAAG

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TTCCAGCCC GAATTTCTAC GGCACGCGA AGACGGAGGT CTTCTTTCTT TCCCTNAGCC 60  
AGCATGCTT COTGTGTCCA AGAAGCTCT GAAGCGGTG GCACTTCCAA AGCATTTGAT 120  
55 GCTGATATA TTGACCGGTG TGTTTGCTC TGTTCATCC ACCGGTCCCC ACAAGTTGAG 180  
AGAGTGTCTC CCCCTCATCA TTTTCTTGAG GAACAGACTT AAGTATGCC TCACAGGAGA 240  
60 TGAAGTAAAG AAGATTTGCA TCCAGCGGTT CATTAAATTC GATGCGAAGG TCCGAACTGA 300

360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1290

TAATACATAC CCTCTGTGAT TCATGATGAT GATCAGCATT GACAGAGAG GAGAGAAATT  
CCCTCTGATC TATGACACCA AAGGTCTGCTT TGTCTGTACAT GATATTACAC CTGAGAGAGC  
CAAGTACAGG TTGTGCAAG TGAAGAAAT CTTTTGTGGGC ACAGAGAGAA TCCCTCATCT  
GTGATCTAT GATCCCCCA CCAATCCGTA CCCCAGATCC CTGATCAGAG TGAATGATAC  
CATTCAGATT GATTAGAGAG CTGACAGAT TACTGATTTC ATCAGATTCC ATTACCCAG  
CGAGTGGTCT TGTGACATCT AGAGCTCTCG GAGATCTCTG CCGAGGCGAG GATGAGAGCA  
AGCTTCAGAG GACTTCAGAG AACTGCTCTT TCTCAGCAG, ATGATCTAC TGTCTTCTTG  
CGAGCTACTG CTTTCTCTCT GTGACACCA CTTTGGGAGAG CCAATTAGAA AGGTGACCTC  
TTGTGGGAAT TCTAGACCA CAGGCGACCA GCTAGAAATC CTGGGCTCTCC TGGCCGCSGG  
GAGACGAGAC CTGCGCTGCA CCGAGAGGAA GCGAGCTGCT ACTGCGAGGC TGAACCTGCG  
GGGAGACTTG CTGTCCGCGC CCCCAGAGAG CTCCGAGAGC CCGAGAGAC CGGGCTCTCTC  
CGCCCCCAG ACCCGCGAGA TCCCTGAGC CAGAGGCGCG GTGCTGTGTC ACCGAGAGAA  
GAGCTGCGG AACATACAT GAGATCTCTT CGGCTCTGCG TTGCGCAGAC GCGAGGCGGC  
ACGAGGAGAC GAGGCGAGAA GCGCTGCGCG GCGCTGAGAG CCGAGGTGCG GCGCATGTAA  
CTTCAGACCC CAGAGAGATC AGAGCATGCG GCGCGCGCGC GCGCGCGCGG GAGTGAAGGC  
TGAAGAGGCG GCGGCTGAGG CTTCAGACC GAGGCAATTA AAGAAATCT GCGTACATCA  
AA

40 (2) INFORMATION FOR SEQ ID NO: 82:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

60  
120  
180  
240  
300  
360

TTTATGTTAT TCTGTACTA TGAAGCTCT ATTAAATCT TTTTGGACT TGTATAGTGG  
TCTTTAAATG TTTTAAATTC CAGCTGAGG TTTTCAATAT TGACTATACC CCGTGAAGAT  
GAGTTGTTTT ATAGACTCTA ATGATGAAA AATCTTACAT CTTTTGTGAG TCTCTTTTCA  
TTTGTCACTG TTTCATGTTGA TTCTAACTCA TGGTAAATTA ATCTGCGTT TTTTTTTTTC  
TGTTAAGAAA CATTCTTTGA AAAATTAATTT GAGAGAAATAT TGAATCTTA TGAAGAGGC  
ATTACTGACC AGAGAAATAT TTTTGTGTTT ACCAGGTGCC TAAAGATGCT AACAGTCTGCG

420  
480  
540  
600  
660  
684

GAGCAGATAG AACACAGAGT GATGAGAGAA TCCCTGAAT CCGTTTTCAC TTGCGGCAT  
CTTTCTTCC AACCTGTGCG GAATATCTAT TCAATATCTA TCTGAGAGCT ABAAGGTGCT  
TTATCAGAGC CGAACTTGGA GGGCTGTGCG CTTTAGTAC TGTACACCCA TCATATCTGA  
GCTTCATGGA TTGAATCTCT TTTTATCTTT CAGATTCTCT TTTAAATATC TTGTCTTTTT  
TTTCTCTCC GAAGATATCC CCGAACATTA CCAATGCCCA CTTTCCGTT AATTTTTTGG  
GCTCTCATTT TGAATTTTTC AAGA

15 (2) INFORMATION FOR SEQ ID NO: 83:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2024 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080

CTGCGAGAT TCGACAGAC TCGCTGTGAG GCTTCAATTA TCGCGCGGCT GCGCTGCGCT  
TCTGGAATTT GAGATCTGGA GCTTTCCTTG TTGCTGTGTC GCGCGGCTTTC GCGCTCTTCT  
GCGCTGCGCG GCTGCGAGGC TGGGAGAGCG GTTGGAGGCG GCTGTTGATC GCGCGCTTTA  
AATTGCGATC GGGCGCGCCA TGTGCGGCTCG CAGATGTAGG GCGCTAGTCT CCGAGCTGAG  
CGCGCGAGCC GAGGGGATTC AGAGAGAGAA GTGCTCTAT GCGCATGAAA ATGAGATTGA  
AAGGCGAGAA GAGGAAATTC CAGTGTCTAA TCTCTCATCT GAGATTGAGG ATGAAATCTCC  
TGAATATGCT GTTACAGAAC CCAATGTGAC TGAAGCGGAA GATGATGAGG ATAGTGCAGG  
CGATGATGAT GAATGATGAG TTGATGTGAC TATGAGAGAC ATTAAAGCG GAGCGCGACA  
GTATGAGAT TATGTGACAG CACTGTAAA TCTTAACATC AAGACAGGCG GAGAGTTTAA  
TGAATCTTACA GGAACAGAAA TGAAGAGAT AAGCTTAT GCGACTGTGAA GCAATTAATCG  
AGTTCACATC TTGAGGTGAG ATTGAGATC TTTTGAATAT AAACATGCG GTAAACTCTGG  
TGTGATCTTT TGTGATTAAT TTAATTAATG GTTAAATGAA GATACCTGGA AAGCTTACTG  
TGAAGAGAAA AAGAGATAC GATGTGATCT TGAATGTATA CCAATTAATCT CTATACAGAA  
TAAATTAATG GTTACGAGAG GAGAGACTCG AATCTGAGG AAAAGAAATC CCGTTTCATC  
TACAAATCT GAGTTTACTT CTCTCTCTTC TTGTCTCAGG ACTGCGCTTC GACCGAGAG  
GAGATTACT GGGCGAATTC ATGTATGCG TGAAGCTAAA ACTATGAGCG GAGTGAAGAG  
CAGGAGAGCG GCAATGTAGA AAGACAGATC AAGAGTCTCT TCTGAAGAGT CTGCTACTGA  
ATGAGAGAC AATTATACA AACCGATCCG GTTTTCTCT CAGAGACTCT CTCCACTTCA



5 CTTTCCAGCT CTTCCATTTT TTCCACCTCC TCCGACTGTC AGCACTGCTC CACCTCTGAT 1140  
TTCCACAGCG GGTTTTCTCT CTCACACAGG CCGTCCAGCT CCATCTCTTA TACCAACAT 1200  
AGAAAGTGA CATTCTCTG GTTATGATAG TCGTTCTGCA CTTCCATTTT CATATGCGAA 1260  
TGTTCCTTTT CCGCATTTT CTTGCTTCTC TCTTTCTGCG CTTAGTCTTG TCGACACAG 1320  
CAAGCACTGG GACTATATG CCGAAGAGA GAAAGACGA GATAGAGAGA GAGACAGAGA 1380  
CAGAGAGGA GACCTGTATC GGGAGAGAGA AAGAGAACCC ACCAGAGAGA GAGAGAGGA 1440  
CGCTGATCAC AGTCTTACAC CAGTGTTTT CAACAGCGAT GAAGAACGAT ACAGATACAG 1500  
CGAATATGCA GAAGAGGTTT ATGAGCTCA CAGAGCAAT CAGAGAAAG AAGAACGACA 1560  
TAGAGAAAGA CGACAGAGG AGAAGAGGA AACAGAGAT AAGTCTTCTC GAAGTATAG 1620  
TAGACCTGCG CATGAAGCTG AAGAGAGGA TAGTCAAGG AGACACAAAC ACAGAAATC 1680  
TAAAGAGCC AAGAGAGGA AAGAGAGGCG CAGTGAAGCT GCGCCCTGAC AGGAGAGCAC 1740  
CGAAGCTACA CCGCAGAAAT AGCATGTTT TTGCGCTTTT GTGTATATTA GTACAGAG 1800  
TAGATATCAT AATCTCTTT ATTTTCTCG ATATGTTTA AGAATTTAC CTTAAATCTT 1860  
GTTCTGTTTG TTATATGCA AAGTAACTT TTTTTCGAA ATAAAGAGT GAATTTTCA 1920  
TGTTAAGTTA AAATCTTTG TCTGTACTA TTTCAAAAT AAAGAGAG CATTGCTTT 1980  
ATATCCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGCC GCC 2024

## (2) INFORMATION FOR SEQ ID NO: 84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 911 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

50 CCGGCGAATA GCGGAGAGGG GATCTGAGCT GCGAGATGA ATGTGGGGGT GGCACACAG 60  
GAATTAAGCC CGACACAGCG AGTGAATGAT AGCGGAGGGA TCTGGCTGGC CTACATATC 120  
TTGGTAGGAT TCGTGCATAT GGTCTTACT AGCATCCCTT TCTTCAAGAT TCTGTGTCT 180  
TGGACCTTGA CCAAGTCAAT CCATACCTTG GCTAGTATG TCTTCTTCA TACGGTGAAA 240  
GGGACAGCCT TTGAGACTCC TGACAGAGA AAGGCTGGCG TACTGACACA CTGGGAGCAA 300  
ATGCACTATG GGTCTCAGTT TACTCTTCC CCGAAGTTCC TCAGCATCTC TCTATTTG 360  
CTCTATCTCC TGGCAGCTT CTATACAGG TATGATGCTG CGCACTTCTT CATACACACA 420

5 GCGTCATTCG TAAGTGTACT GCTGCGGAG TTCCCGCACT TCCATGGGCT TCGTGTCTTT 480  
GGCATCAACA AATACTGAGG GATGGGTTTT GGGACAGCTC CATGGGCAAT GGGAGAGCAC 540  
TGAACAGAG GACTATTAATA CATCTCTCTC TTATCTTCCA TACTGTCTTC TACACTTTTA 600  
AAGCTGAGA ACTATACAC CTTTCCAGA CTCCAGAA GAGAAGAT TCCAAATGG 660  
GGCTCTGGG CCCAGTCTTG CTAGTGGCAA GTTTCTTTGA ATCAGAGAG CAGGTGAGST 720  
AAGGCGCAA TCACTCTCTT CATTAGAGG AAGCAATTTG GCGAGCTCTT TTGCTGATTA 780  
CATCTTCCA TATCTTTTAC ACTTACACC TTCCAGCTCT GTTTTGTCTT GTATTTTCTT 840  
TACATAAAT TTTTTCAGCT ATAGCTGCGG TTTAAACAGG ATGGGTAGAG AGCTGTCTTC 900  
ATAGGCTGG GGTGGGAG ATGAAATACT G 931

## (2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 825 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CCGGCGGCGG GGGGTCTTCA GGTACCGGG CTGGTTACAG CAGCTCTTACC CTTCAACAGG 60  
CAAACTGGC AGCGCAGAG GACACAGAGA AAGATCCGA GCGGAGGG CTGAGCGGCA 120  
CGACCTGCT CCGGAGAGCTG ATTCCCTCCG GTGAGGCGG GAGTGGCTG GAGCGGCGCC 180  
GCGCAACAT CCGGCGCTGG AGCACTTTCG TGAACACGA GCGCTTCTCA CGGCGCGGCA 240  
ACCTGGAGA GCTGTGCCAG GCGCTGTAC GCAACTGGA GTACTACAG AGCACTATG 300  
TGTTCGTGT CTTGGGCTC ATCTGTACT GTGTGTGAC GTCCCTATG TTGCTGTGCG 360  
CTTGGCTGT CTTTTCGGC GCTGTGACA TTCTCTTACT GCGCACTTG GAGTCAAGC 420  
TTGTCTCTT TCGCGAGAG GTAGGCCAG CCAATCATA TCTCTGCTT GAGGCAATCT 480  
CCTTCCCTT CTTCTGGCTG GCTGTGGCG GCTGGGCGCT CTTCTGGTG CTGGGAGGCA 540  
CCCTGTGCT CATCGCTCC CAGCTGCTT TCCACAGAT TGAAGCTGTG CAGCGGAGG 600  
AGCTCAGAT GGAACCGTG TGAAGTGTCT TGTGGAGCTT CCGGCGCTCC GCGGCGAGCT 660  
GCGGCAAGCC TCGGCATGCC TGTCTGTGAC GGTCTGTGCT CTGGGCGCA CAGCGCGCTC 720  
CCATCAAGC CCGGAGAGG GATCCGCTT TTGAATAAA AGCTGTATG GGTGTGATTC 780  
AGAAAAAA AAAAAAAG GGGGCGCTTC TAGGGGTCAA AGTTA 825

## (2) INFORMATION FOR SEQ ID NO: 86:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1238 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

5 CAGTGAAGG GATGAAATG GACTCTGCT GTTTTATT TTCTATGAG GAGCTTCTG  
GGAGAGTTT CTGGCTTCA GACTCTGAGA AGCTGCAAT TTATGATGAC TCTCTCTG  
15 CTGCACATA CTGGAGAAC CATAGCTGC AGCTTTAGA AAAAGAAC CCGGCGAGAG  
TCTGGGAGAG TGGATGACA GCAATGCAAG GCTTTGAAA ATGAGAGTG AGAGTATTC  
20 CAGGAGAGT GTAGAGAGAG GATGATCTT GATACATGA TTGAGATCA TTAGGCTCT  
GTCTGGACA CTGGCTTCC TCGTTACTG CTCTTCTCT CCGCTTGGT CGAGAGAGG  
25 GCTGCTCAC TCGTCTGCT TCAATTTCA GAGCTGCTG CTGCAGTCA ACTTAACTCA  
TCTCTCTCA CTCTTCTCT TTTGCAAT AGTGAAGTG ACAAGATGT GAATGGGCA  
30 GGAATGCTT TTGATGCAAT CAAAGCTTAA GCTTCTGAG CAGTGTGCT CAGCTGAT  
TTGAGTTGCA GCGTGAATG AAAATTTACA TGGAGCTGA GACTCTCAT GAGAGTAAA  
35 TCCCAAGTT TTTCTCTGT CTGACAAAG AAACAGAG TAACTGATT GCGCTGAC  
CTGGCAGTT GCAATTTCCC TGCAGCTTG ACCCAAGCC AAACCTGTA AAAAGTATC  
40 AGATTGTGC CAAAACACT GAAAAAACT GCGCTGCCC TGAACCAAT ACCTGAAAC  
CTGTAAACT CAAAGCTG ACCCGCTGT TTTGATATA CCGAGTGA ACACTCTCT  
45 CTCACTGCT GTTGTAGGA TACCTGTAG CCACTCATT AATGACATC TCTTAAATA  
TCTTTGAGC TATGACCTT GCAATCTTT TGTCTGGG AATCTAACT TTATCTGAA  
50 GTTCCAGAG GCACTGATA GGAATTAAC ATACTTTAA TGGCTTCTCT CTCTCTGTT  
TTAGCTATG CCGTCACTC CTGAGTTAAC CTCCAAATA CAGATTCAC CTGTAACAA  
GCGCTTAAT TGAAGATAC AGATCACT GTAAAGAAC CTGTAAGTA AGCTCTCT  
TGGAGAAC GAACTTAAA CAGTCTCT GTCCTCT 1238

## (2) INFORMATION FOR SEQ ID NO: 87:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1460 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

5 ATTGCTCTCT GATCCCTGCT GACACTGGG TCAATCTCA TCCCGGAGA GCAATTTCTG  
10 CTGCTCTCT TAACTCGGG CTGATGGG GTGCGAGAG CAAATTAATC GACATCTGCG  
CCCACTCTA TTGCGAGACT CTGATGAGC GACAGAGCG ACCGATCTT CAGCATCTTC  
15 TACTTTGCA TTCCGATGAG CAGTGTCTG GACTTAATG CAGCTTACA AGTAAAGAT  
ATGCTGAGAG ACTGCACTG GACTTGAAG GTACACTCG GTCTAAAGT GATGAGCTTT  
20 CTGCTCTGT TCGTGTAAAT GCGGAGACCG CAAGGGGAG CCGTGAAGCG CCACTCAAT  
TTGCAACCC TGAACCCAC CTCTGATGAG GCAATCTGA GCGCTTGGC AAAAATCTT  
25 AGTTTCTCC TGTCTTCTT GGAATTCAT GCTGTGACT TTGTCAAGG CTGCTGCTT  
CTGTGAGCTC CGGATCTCT GATGCTTCC CCGATGATCC TTGGGAGAGC CCAACCTGC  
30 ACCGAGTCC TGGGTGTGAG CTTGATGAG GAATGACCC GCGGCTGCG CCACTTCAC  
CCCGGAGTGT ATCCGATGT CTGTGCACT GCGCTCTGAG GATGTGAC CTGCTCTTC  
35 CTGTCCCTG CTGAGGCCC TGGTAACTC GTGACACTT AATATTTCAT CTTCATTGA  
GAGACCTTCC TATTCATGA CTGGGCAATC GTGCGAGCA TTCTGCTGA CCGTGTAACT  
40 CTAACTGAC GATTCACGCG CAGAGCTTC GAGATGATC TGTCCACCT GCTGGGTGAT  
GATGAGAGCC CTAACTCAT TGGCTGAATC ATGACACCG TGCACAGAA CTGACCCCC  
45 TCTTCTTGT CGAATTTCCG GACTTGAAG TTCTGCTCA TCTCTGAGC GTTGTGTGAG  
GCACTGAGG GCGCACTTCC TGGGACGCC CATTCTCAAT GAGGGGAGAC GCGGGCGGCG  
ACAGCTGAC GTGAGAGGCC TCTTACAGA AGCAAGTTC ACAAGACAC GAAATTTGCT  
50 GCGGAGAG GCGGCTTCA CCGGCTGAC GATGAGAGT GTGCTAATC GAGAGAGCTG  
GCGTCACTT ACTGCAAT CTGCAGAGC TGGCTCTGAG CCAACCCAC GAAAGGCTG  
GCGTAAAC CATTGAGCG GCGGAGTTC CAGAGGAC CTGAGAGTGT TGCAGCTTCC  
55 CAGACATAC ATGATTAAT CAGGAGAGA GATGGAGATC CAGAGAGGAG ATCCCTTCC  
AAAGAGGGA GCGGCAAGG CTGATCTA TTTGTAAAG GATTAATTT TATGACAGA  
AAAAAATA AAAAAAATA 1460

## (2) INFORMATION FOR SEQ ID NO: 88:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1395 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

5 CAGGTGCAAA GTGGAGATG TGAATCTCTA GTCTGGGCT ATTGGCCAC GTGCTCTGG 60  
GACATGGGAC CTGGAGGCT CAGCAGCTG GAGTCTTGGC CTTTGGGTC CAGGCTGG 120  
15 AATTGGCCA TTGCAAGCC GGAAGCTGG ACTCAGGCTG CCCCAGGCT GTTCTCTATC 180  
CTTCACCGG ATCTGTGGC GTTGGACTG GGGCTGATG AGTTTCTGA CTTCTGACC 240  
GTAATCTCTC CAGTTTAAG GTAGACATTT TGGAGGCCC AGCCAGAAAC GTCAACCGGA 300  
20 GAAAGTTCAC CCGGCGAGAG CGKCCCGCT GTGTGCTGCC CCGGAGGAC AGCCAGCTTG 360  
TAGGGGGAG TCCACCTGA AAAAAAATT TCCAGTCCC CAAAGGTGA CCGTCTTGG 420  
25 GAGACGGG ATGCACTACC ATGTGGGTGC CCACAAAAT TTCACCTTG AGTCTCTAC 480  
TCTGACCCC GGGTCACTT CCAGAGCAA GBACTCCCTC CTGCTTGGAA GAGACTTCAC 540  
30 ACCGTATCA CBAATCCAC GGTCTGAAG GTGGATGGA TTCTTGGTG GATTCATCAC 600  
TCCCGATCA AAAAAAGCAA CBAAGCCCAA CTAGAAACAT GGGTCCCGAG GGTGGGTCA 660  
GGCCCTTAA ACTGCACT AGTTGGGTG AGCCATTAG ATTAATCTT TTCTTAAT 720  
35 TTGTAAACA ATCATAGCT TCTGTCACT TATGTATCTT AAGACTCAAT ATAAACCTT 780  
TGTATATCT GAGGAAATCA ATGATTTGAT TCCCAGAAA CAGAGTGGG GAGTGTAGT 840  
40 TCCAACTGG TTTTACTTAA CCGTGTCTT AGACTCTCC TTCTCTTAA TCACTGAGCC 900  
TTGTTTCCAC CTGAATGAC TCTCCCTTAG CTAAAGGCG CAGATGCACT CCATCTTGG 960  
TCTTTTACT GGCAGGCGCT TCTCTAAGG ACTTAACTTG TCAAGTGA CTCCAGGAC 1020  
45 ATCCAGAT GCATTAACT GATAGATAC TGTGGCAGC TATATCCGA GTTCCAGGA 1080  
ATTGTGCAA TTGATTAAC CCAAAAGCCC GCGCTATC ACCTGTATC AATCTTAAG 1140  
CCCTGCACC TGGAACTATT AAGCTTCTG TAAACATTA TCTTTTAACT TTTTGTGCT 1200  
50 ACTTTATTT TGTAAATTC TTTTAACTAG ACCCCCTTC TCTTTCTTA ACCAAGTAT 1260  
AAAGCAAT CTAGCCCTT CTTACGGCG AGAGATTTT GAGCTTAGC CCGTCTCTGG 1320  
CCACAGCTA AATTAACGA TTCTCTATG GTAAAAAAA AAAAAAAA CTGCGAGGG 1380  
GGCCCGGTA CCCAA 1395

## (2) INFORMATION FOR SEQ ID NO: 89:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1186 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

5 GGCACAGCC GGCAGCCGA GCTAGGTCG AACTGGGG GGCACAGGA TGTNACAG 60  
AAAGCAGAA GATGAGACTC TTTTCAITCA CTTTCTCTAG CCCATCTTG TGTCTATCTT 120  
15 TCCCCCTCC ATCATACCTC CTCCTTCTG GAGCCTCTGC GGGTGGCT GTAACTGGG 180  
CACTTACCTG GATATTTCAG TGGAGGATG AAGGCGAGA CTCACCTTAC GCGTGGGAC 240  
20 AGATGGGAG AGGAAAGAG CAGAGATGC CAGGAGAGG GTCCAGACA AACGAGAG 300  
GTTCGGTCA GGGAAAGGG TGGGAGAAA GAGGGGTGA GGGCTTGGAG GCGGTTAGC 360  
CAGCAGCTC GGGCTCCCG GGGCTTGGC ATCCAACTTC GAGAGAGGG TACGAGCTC 420  
25 CTGTGTGTA TCAATAGGAT TTTTACATA GTGTATGCA TGAATCTCT AGGTTAAGA 480  
AGCGGTGTG GTCCACCATG ACATCCACC GGTATATATA AAGATAAATA TATATATATA 540  
30 TGTATCTAA TTATGGCAG AGAATATATA GCACGTAGGG CCGTCTGCC CTGCTGAGC 600  
AAGCAAACT AAGCCTTTTG GTTGGGTAT TATGTTTCT TTTGTATTT GTTGTTTTT 660  
35 GTGCTCTTC TTATGTCTG ATAGCAGAG TCCAGTGGG ATTCTCTCT ATACAGAA 720  
AGTGTTTTA ATTCTCAT GTTCTAGTTA ATGTCTACT CAGCAGCTC TCTTAGCTTA 780  
ATTTAGGAG GTTCCCAAT TTTGTTCTT CAATTTTACT GGTACTTTT TTGTACAAAT 840  
40 CAATCTTTT CTCTCTTCT CTCTTCCCA CTTCTCACC TTGCCCCCTC CATCTCCCTC 900  
TCCGCGCTC CCGTCTCTC TCTGCTGCC CTTCTCATTT CTGTCACTC CATCTCTCT 960  
CCCTCTCTC TCCGCTCTC TCCCGCTCC CAGCGCCACT TCCCGAGTT GTGCTTCCG 1020  
45 CTCTTATCT GTTCTAGTTC CCAAGCAGTT TCACTCGAAG TTGTCCAGT CTGCTTCCAG 1080  
CTTTCGAT CTGCTCTCT TTCTGTAGA TTGAGGCTT TCTTTGTAT TTCACTGTTT 1140  
50 CTGACAGT TTAAAAAAA AAAAAAGAAA AAAAAAAA AAAAAA 1186

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1821 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

## (2) INFORMATION FOR SEQ ID NO: 90:

311

(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

5 AAAAAATGCT TTGAGGGCGT CCGGTATGTA TTGCGGGGGGCG CAGAGAGAC TGAAGCTGAA 60  
KATCCGTGCT GACTGGCTC AAGATGGCTT GAGGAGAC GAGTTACCA GGTGAAATTC 120  
ACATGCCCTC CAGAAATCT CCGAAGCTA GTTGTGAGCA GAGGTGGGGC TGTGAGGCTA 180  
GAGTTGAC ACACATGAC AAGCTCTGCG GGGACATGCG AGAAGCCCTT TTCTTTGCT 240  
CTCCATGCTA GAGGCGAGCC TTAGAGAGCT GAGTTACTC CCTTGAGTC AGCAGAGAGC 300  
TGAATGACA GAGTGGCCCA CAGAAAGCA GTTGTATGAC ATCTGTTTAC GTTGTCCGAT 360  
GAGCGAGCA CAGGCGCCCT TTCCAGGTCA AACGAGAGC GATGCAATCT CCAATTTGAC 420  
AGTCAATGCA CTACTTCAT TGTACTGTTA GATGAGCAA AGTGGTACT GCTGAGTGT 480  
GTGCGATGCC GTAGAGGCGC GAGTCAGAGG ATGCCCAAG TCCAGGCGCTG 540  
GAGGAGAGC AACCCGACA GGTGATGTC AGCAGAGAGC GTCCGAGCCA ATGTCTGTGT 600  
CTAATTTGAA GGGGAGTTA GCGCAGAAA ATCAACTGAT GTATATTTAC AAAATTCCT 660  
TGACTGACT TAAATGCTT ATGAGACTT ACTGCTTTTA AACCTATTTG AAAATCTGTA 720  
TACTTAAACA AATGAGCTC TTAGAAATTT TGGAGAGCCC ATGCTAGAAA CAAATTCGT 780  
GAGTGAAGAA TGTTCGAGA AGTACTTAC CTACATGTA ATGTGCAAT TTCTTCCCT 840  
TTGTAGAAA AATCCCTTT ACTTTTGA ACGATATGCG GAGTCTGAG TGAAGCAAT 900  
ACATGTCAT ATGAGAAAA TTAGAAATTA GTATAAATTT GTTAAATAT CAAATATGCG 960  
ATATGATGCG CAGAGAGATT TAACTTACA AGTATTTCT AATGTAAAG TATTCAGCTC 1020  
TTCCAGGTG AATCCCGCTT AACCGAGCT GAGCTTCAC ATCATTTCA AAAAGCAAT 1080  
TCTGTGTC CAGATATCT CCAATAGCT AACTTTTGG TTCTCGATTT AGCAATTT 1140  
AAATCTGCA AAGATTAAT ATCATGAT TATTTAGCT TTAGAAAAA AAAATGTAT 1200  
GAGCTTTAT TTTCATGGCC TATAGAGAG TACCTTACT GGGCAGATAT AGGAAAAACA 1260  
AATTAAGCA AAAAGAAATC TGTACAATC CAGAGAGAG AAAATGTGCG CAGATGACT 1320  
CAATTCGCG TCCCTCCAT CAGTCAAT CAGAGAGCTG GATGATGCG CTGTCTTTG 1380  
AATGTGAGC AGTGTGCTT GTAGCTTTT AAAATGGCC TATAGAGCT TTAGACAGAT 1440  
ACATTTAAA GATGAGTTA CTTAAAGAGC AAAATTAATTT TATAGCTTT CAGCAATTTA 1500  
TGAATGTTG GTTGTCAAT TTTTTCAAT AACTTTTCT ATCAAGTCT GTTGTCTTTG 1560  
TAAAGATTC TCAATTTGG GTTGAAGAG TAAAGACAAA CTGTATTTTC AGTGAAGAAA 1620  
GTATGTTTA TGAATTTTA AACCAATTA ATGTATAAA GGTATAAAAA AAAAAAAA 1680

AAAAAAAAA AAAAAAAAAA AAAAAAAAAA CTTGGGGCGC GAGCTTTTCT CCGTTTGGCT 1740  
GAGGCTTAT TTATGCTTG GAGCTGGCG CTTTCTTTT TACAGATGCG TGAAGGGGGG 1800  
5 AACCGGGGGG GAGTTTCCCC C 1821

(2) INFORMATION FOR SEQ ID NO: 91:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

20 TCCCTTTT CCGACGATTT CCGGGGATG TGAAGGTGCG AATGTGAC TCAATTTAG 60  
GAGTGAAG GAGTGAAG AATTTGAG TGGAGATTT GATTTGAT TGGACTGCT 120  
AATGACTTG AACTGCAAT CTGTGTCA GGTGAGGTT TGTGTGGG TTCTTGGAG 180  
AGCTTACTA CCCCCGATC TTTTCTTCT CTGTGTCA GAGAGCCCT GTTGTGCTT 240  
TACAGGACT TGAATGCTC CAGAGACA AACAGGAGA GAGGAGCTG TAGGAGAGT 300  
TCTTCTGCT TTTCTGCTT TCCCTTTTA CAGAGCTCC GAAAGGAC TCAATGCTAT 360  
GCGAGAGCT TTCTGAGAA GATGATTA GATCTGTC AGTCTTTTC TTGTCTGCC 420  
AGTGAAGTT TGTATGCA CCGTTCTGCG TGTCTATGAG AATGATGAG AAGAGCTGCG 480  
GAGTTTTC TGTCTCTC TGGCTCTCA TGAAGGATG GAGCTGGGCG TCGGGGCTC 540  
CTGAGCTCA GATTTATTT CTTCTCCCG TCCAGCCCT TCTTTTGT CTGAAGCGG 600  
TTTAAATG TGAATGCC AAGAGAGAG CCGCTGCTG TATGAATTT GACGGGCTTT 660  
TTGTAGGTG CAGCCCCA ACTTGAAGT ACTTAAACA ATTTTAAA GATTCATGCG 720  
CTGTGATC CTGAGAGT AGTATGAT GTAGCTGCG CAGAGAGT TGTGAATTTG 780  
45 TGTGCTGCT GATTTATGAG ATTTCATAT TAAATGCTT GCTGATTA AAAAAAAA 840  
AAAAACTGA GGGGGGCGCG GT 862

(2) INFORMATION FOR SEQ ID NO: 92:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 696 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

5 CTGAGGGCAG TGAAGTGGAC TCTGAGGGCT ACCCTGACCG CCAGTGGCTGC GCGAGGGGGG 60  
 TGGAGGGCAG AGGGGGGGGG AGGGGGGGGT TGCAGAGATG GCTCAGAGCA GAGAGGGGGG 120  
 AAACCGGTTG CCGGAGCCCA GCGAGCTTTG CAACCCCTTT CAGGAGCCAG CTGTGATCCA 180  
 GACCGACCC AGCGGGCAGT ATGCCAGCTT TGAAGTCTAC AACCTTTTGG AGAGCCGGGA 240  
 GGCACACCA GCTATGAGC CTTCAGGCC CTGCCCAATG CTTCCACCTT CAGCTGCCCTC 300  
 CTTTCAGCCC TCGAGAAAGC TGAAGCCCGAC AGAAGCTTAG AACTATGGCT CATAGAGCAC 360  
 15 TCAAGCTCTA CTTTCAGCAG CCAAGCTGTA GCTGCTGAG AACAGAGAGG AGCTCAAGCG 420  
 GAAGCAGAG CAGTTGGACC GAAGAGCGA GAGCTGAGC ATGCTGCCCT GGGGGCACA 480  
 GCTACTCGAC AGAACAAATG GCGCCCTCTA CTTCTTTTGT GTCCAGTTCA GCGCTGCTTT 540  
 20 TTCCAGACA TCTCCATGGA GATCCCGCA GATTTTCAGA AGACTGTATC CAGCATATAC 600  
 TACTCTGGA TGTGAGCAC GGTGGTCTT CTCTGAAAT TGTGCGGCTG CTGCGGCAAT 660  
 25 TCTGTGGA AACCAAAAT GCGAGGCTT TGGGTT  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1886 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (2) INFORMATION FOR SEQ ID NO: 91:  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:  
 40 CAGGCGACTG AGGCTTCTTT CCGAGGGATG CAGGAGTCC TACAGAGAA GCGGCTTCTT 60  
 GCATTTGAGA GGGGGCAGAG CTTGTACCC AGAGATGACC AAGCAGCTTT CTAGCTGGCT 120  
 45 CTGAGCTTG CCATCTCAG AGAGTCCA GAGGCTCTGG GGTATGTGCG CCAGCTCTT 180  
 CAGCTTCAG GTGAGATGC CAATCTCTG CAGCTCTTGG CCTCTCTCTT GTACGACAG 240  
 AAGCATTACC ATGAGGCTCT GACATCATC GACATGGGCC TGAATGATA CCGAGAAAT 300  
 50 TTCTACTAC TGTTTTCCA AGTGAAGTG CAGTCACTCT GCGAGGGCCC GAGAGAGCA 360  
 CTCTGACTT GTAGACAT GCTCGAGTA TGGAAATCT GCTACACTT CAGCAAGCC 420  
 55 AGTGAATCT GAGCTGGAG CAGGCTCTTA GATAGAGCA TTGCTGACAG ACCAGGCTT 480  
 AATACAAATTA CTTTTCAGA CTTACAGAT CCGAGAGAG GCTTCGTTCA TCGCAATG 540  
 GTAGCAGGCTT CAGAGTGA GCGGCGCTG TCGAGATGG CTTTCTCTCT GCGAGAGCTG 600  
 60 CCTTACGA GGGCGGCTG CAGCCCTGGA TGAAGCTGGC ACAGATCTGC CTCCATGAG 660

CTGAGTCTA THTGGGATC GCGAGGCTG CAGAGGCCAC AGCCTGTACC CAGAGAGCTG 720  
 CCAGCTCTT CCCATGTCC CACATGTGTC TCTATATGCG GGGCCAGATT GCTGAGCTCC 780  
 5 GGGGAAGCAT GAGAGAGGCG GCGGGGTGGT ATGAAGAGGC CTTAGCCANT CAGGCCACAC 840  
 CAGGTGAGA GATGCGAGCG ACTTGGGCTT GATCTTAC CAGTTAGGCC GTTACAGTAT 900  
 GGGCGAGAG ATCTTCGGG AGCGGTGCA GGTCAACTG ACAGCCACAG AGGTCTGGA 960  
 CCGGCTGGGC GAGGTCTCTC AAGCTCAGG CAGAGATGCG GCGGCTACCG AGTCTTCTCT 1020  
 GACAGCTTG GAGCTGAGG CAGAGAGCTT CCGGCTGCG TTACACATCA TCCGCCGCTT 1080  
 GCTCTGAGCA GGGGCTGCC AGCTCACTT CCGGCTGAG CTNCGAGGC CTTGCGGGC 1140  
 ACCAGGCTT GTCCATGCG CCGAGGGGA TGAATCTGCC GCACTGAGGC CAGGAGAGAG 1200  
 TGTTCAGTGG GGCACAGTGA ACCAGCCAA CCAACCCGGA ATCATGCTC TCGCCATGTTG 1260  
 CGTTCTCTT GTTTTCTTG CAGGCCAAT GTTATGTTCT GAACTATTTG ACATGTTTCA 1320  
 20 AATGATCA TGTGCCATAT TTTGTTAGTT GACATCTGAG TTTTCACTAA AATGATTTATG 1380  
 GATTTATCA GCAATGTAG AAGATATAT TCAAGTTTAA AATTCAGTGG CAGACAGAT 1440  
 TATTTTATC AGAGCTGTAA AGAAACAC TCCTCTTTC TCCCCACAC CCTCTCTGCC 1500  
 CCACTTTGCC CAGAAACA AATGTGAAT TCCTGTCTCC CACCTCAGCA CTAGTCCATG 1560  
 CAGGACACC AGCTGACAT TTTTGTGTTT TACTGTCAAT AATTTGACA TGTGATCAAT 1620  
 TACTCTCTC ACTTAGACA AGGCTGAGT CCGAGATAT TATTTTITA CCAATATATG 1680  
 35 CTTGTACAA GAGAGGAAA TATGATTTT TTAAGTTTAA CTTTTTATG TGAATTCAGA 1740  
 GTTATTTAT CAGGGAAT ATGTACAAAG AAGCTTCAA TGAATATTT ACCGACATTC 1800  
 CTTATACATG ACAGACACT CCGTACATGG GAGATGATG TTATATATTA AATGATTTTT 1860  
 AATCGAANA AAAAAAAAAA AAAAAA 1886

(2) INFORMATION FOR SEQ ID NO: 94:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1774 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

CTGAGCTACC GTATACAGTA GAGATTAACC CAATTTGACA TGCATACAC TGAAGACTTGC 60  
 CTCCTCTGCC CCCCAATTGA AGATGTTCTT TTTTCATTAAC TATATACTAT TCCATTCGAT 120

GAATATCTCG TAAATTAATT AATCCCTGAT GAAATGATA TTAAGTCAAT TATATATAGA 180  
AGTATATATA AAGATCTGCT AAGATATATT TGCATCTGCT GAGGATTAATT CTATAGAGAG 240  
AATATACAGA AATTAATGAG ATCAGATTTT AATATTCAG TTATGAAAC TATACACAAA 300  
AAGATTCAG CAATTAACA CTCATCAATT AGTAAAGAG CCGTTTGC CATTAGCTGC 360  
CAACCTGAA TCCCTAATAA TTTTGGGAAA TCGATAGGC AAAATTTCTT TCTTTTCTTT 420  
GAATATTAAT GAGAGGAGC ATCTTCAAT GTTTCCTGGC CATTGCAT TCCATTAATG 480  
AATTCCTTT GCGCATTTTC CTTTTATA TTATGAAAT GTAAATACA CCTTCATAT 540  
GTATATATTA GAGATCTTT TGAATAGAG GAGCTTTTC TCGAATGTA CCAATACAT 600  
TCCATTTAC AAGTTTAC ATACATCTTC TACATCTT TCCATAGAA TATACATACA 660  
CAATATACA CTTTTACAT AATATGATC TCAATCTCG TACCTTTTA AATTTTGGT 720  
CTCAAAAAA GATACAGCT CTTTAAATT CTTTAAATG TCAATATGAT TAAATATCAT 780  
GAATATGCA TTAATTAATC CTTAAATTT TTTCCTGCG CTATACAT GCGAAATGTA 840  
AATCTTAAT CAATCTTT TGTCAATGAG TGTAAATAT TCTTAATCT GAGTCCAGT 900  
AAGATGAT TTTGATACA AAGATTTGTT CTCTGAC CTTACAGCTT CCGAAAGGCT 960  
TTCAATAGT TATTTACAG AAGTATAG AGAATGTA TTCCCATTA TACATACAG 1020  
TACATCTGAT AATTTTATC TGTGAGGCA AAAAGACCT TTTCATATT TCAATTTCC 1080  
TCAATTAATA AAAAATAGT GAGATTTGGG TTATTTCAAT GTTAATTTGC CATTATATAT 1140  
TACATGTA TGTTTGAT CCGTTACAG CTTCATATG GATTATGCT GAAATATAG 1200  
TTTATATTT TTCAGATCT CCTTCCCAT CTTCATGTA CAATACCTT AATTATTTCT 1260  
GGGAACTTA TTCCCTGCG CTAGATAG CATATAGCA GCGCTGGGCT CCGAGTCCC 1320  
AAGATCTCT AAGCAAGTGT AATAAAGAT GGTATATTA CTTAAGCAG GCTAAGGAAA 1380  
GCGCTTACA GAACTTCAG TGCATACT GGAATAGAG CTTTAGCAG ATCCAGAAA 1440  
CGAAGACA TGTAGCCTG AATTTGCTC ATGTAGAG AGTCTGCTG AGAAGAACT 1500  
CGAATGAG CAAGATGGA AAGAGACTA AGTCTGATG TCAATTTCT GAGAGCCCTA 1560  
GATCCAGTG TGCCTAAGC CTGCGCTACT CCGAATTTA AAGTTTGTG AAGCAATAAA 1620  
GTCCCTTTCT TGTTAATAT AATTAATG AGTTTCTGCT CTGATTAATA TACATTAATT 1680  
GTATTTTCTT AATGATTTCT AAAAACTT TGTATTTTA AATTCAGAG TTATGACAT 1740  
AATTAATGTA AATAAATGAG CATGCGCTTC CAGG 1774

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2503 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

GGCAAGAG AGGCAAGG GAGACAGT CAGAGTCCA TCTGCTGCC ATTTCCTTC 60  
CACTCTCT TTCTGATC TGAATAGA AAGCCAGCA GAGAGAAAT TCAAAAGAC 120  
AACCTAAT TCTCTCTCT CATTTCAG AGTTTCCTG TGTCTCTGC ACCTCTTTC 180  
TGTCCCGCG CAGAGGCGAG TGAAGATGGC GAGCCGAGAG CCTGCTGAC CAGACAGCT 240  
GCTTTGATG AGCTATAG TCTCTGAT TGTGTATC TCCATATCT TATACCTCT 300  
TCTGTGAG GCTGAGAC TCACTGACT GCGCAAGTG AATATGGCT TGTATACAT 360  
CTGCTGTAG AATGAGACA CAGACACT AAGATTCAG CAGTTCCCTG AGCTGAGAC 420  
CTGAGGATG CTTGAGATG GCTGAGCTT GAGCAGGCTT GAGCTGAG GGTCCCTGCT 480  
CTGACCTCT TTTCCTCCCC ACCTCTCTT CATTAGCCAG TCGAAGATAG ATGAGAGAG 540  
GTGAGGCTG GAGATGGCT TCTGCTGCT ATCTCTGAG CTGCTGAGAG GCGGCTGAG 600  
CCTCTCTCT TCTATATGAT GGAATGGTTC AAGCTTCCC TCCGAGGACC TGGATTCTA 660  
GCTGTGAG GCGCCAGAGC CTATATATC CTCTTCTTA TACGATAGAG TGTGTTCCCT 720  
CTGAGGCTG AAGAGGCTGA GAGCAGCTT GAGAGCTCT AAGGCTTAC GTGATTGCA 780  
GATTTCAAT CCAACATAG TCAAGATGAG CAGATCTGCT GAGCAATTC AATTATAGC 840  
TACATCTGAT CTGACACTCC AAGATAGAA CCGATCTAG AGAAGATGAG GCGCTGTCT 900  
CAAAAGGCT GAGGAGCAGC AAGGAGAGC AGGCAAGCTG TCACTCTTA GTACAAAT 960  
GTCTGCTCT CAGAGCTTCC AAGCTGCCA AAGATCTCT AAGCAATGCA GCTCATGTC 1020  
AAGCAATTC CTGCTTTAT TATATATCT GAGCAATCT TCCCTATCT TCAAGAGGCT 1080  
GGGAGAGAG TGAATCTGCT CATGGAGCA GACTTCAGG CTGATTTGCC AATATGAAA 1140  
AATGAATTA GCAAAAGACT TACGAGACA AAGAGAGCA TTAAGAGAGC GAGACCTCA 1200  
GTGTCTGAG GAGATATGTT AAGAGCTTC CACTGACCCC ACGATATGTA GTGAGCCCC 1260  
AATAGCATC ACTGAACTC CAACCCAGA GATCCAGAG TGAATCTGA GTGACTTAC 1320  
AAGAGAGCA CAGATGGGCT ACAAAGACA GCTTGAGCT CTTCAAAGCT TCCCTGAGC 1380  
TGAAGCAGA CAGGAGAGAG GGTGCTGCTG ACAAATCACT CCGATATGTA GAGCTGAG 1440  
GATCCAAAT CCGCTGCTG AAGAGAGCTG GAGCTGTGAG CAGAAACAA GCTTGCAGC 1500  
CTGCATCTCC CAGCCAGAA CTTCAGAAA CATTAAGCA ACTTACGAT TGGGATTTG 1560

5 TGTGTATTTT CTAGCACTTG TGTATTGGAA AACTGTATG CAGTGAATTT ATTCAATAT 1620  
TCTGTTCGAA ACCGACACTG AAMACAGAG CAGACAGATG TACTCTGGTG TGATCTCTTG 1680  
TCTCTAGTGT CTCTCTGGG CTCTGTCTCC TCTTCTTTTA TACTAGCTG CCGGGGAC 1740  
AAGGTACAGG TGAAGGAGG GTAGAGCTT GCGGGAGGAG CCGTGTCTG GTTACAGCTC 1800  
10 TATACACTGT GCGCTCAAC TCCACAGAG GGCAGAGAC TGTGGGAGG TCGTTTCTTT 1860  
TCTAGGCTGG CTGGAGAGGT GCGAGCTCAT TGTAGACTC ATGATGGAAA CTATTTTGA 1920  
AACAGGCTTC CTCTTCTAGG AGAGATCATG CGGACTTAAC TTATGCAATT CAGTGCACC 1980  
15 TGGAGTGAAT CTTTTTCTTT GCAAGTACT GTCTCTTTGG TTCCAGTAG TTGGACCAAC 2040  
ACATGACATY ATTTTCTCTG GAACTGGTC ACTGACTTAC ACAGACATTT GCGACTCCAG 2100  
20 AGCCTCAGG GCGAGGAGG GGCACAGTAC ATACAGAGG AGTCAAAATG GATCTCATTT 2160  
TGAGTCTCTG CTTCGCGACA CTCAGACGG CANCCGCAAG GCGCGAGTG TCCAGGGCTT 2220  
25 CTGGCTCAG GTGATCTGC CAGGCGCAAG AAGGCAJAA GTTAGAGCA CAGAGGCC 2280  
CATTCGACA GCGGCGGGC CCGAGCAGAC CAGTGGAGC TCAGCTGTCC TCCAGCTGCT 2340  
CTCGGAGAC AGTTAGTGC ACAGTTATG CCTAGCTGA AAGATCTC CCGAGGTAT 2400  
30 TTACGACAT CCGTTCTCTC CTCTCTCA GGGCTCTGC TACAGGCGA GCTGGAGCC 2460  
CCGCGCTCT GCGAGGGCT GAGGCTGGA GYCATGCTT GTC 2503

(2) INFORMATION FOR SEQ ID NO: 96:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2801 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

5 CTGGAAGGCC GAGGTAGCC GAGCGGGGG GCGGCTCTGG AGCGCGGGT GCTCGGGGTG 60  
CGGTCCGCTC CGCCAGAGC ACCGAGAGC CGAGCGGGG CCGCGCGGCC TCGTCTCTCA 120  
TGAGGCGCGA GTGAGCGGG GCGGCTTAG CCGACCGCG GCGGCTTCCC CCGCGCTCT 180  
ATCGGAGCG CAGGAGAGC GCGGCTTCA GAGGAGGGG GAGGAGAGG AGCTGTGGG 240  
55 AGGTTTCA TCGGCGCCA GTGTCTGTC GGGCGCGAR CCGGCGGCTG GCGCGCGAG 300  
GGGCGGGCC TRAGGCGGC GGTTTYCGA AMGARGGR GCTCTCTAG CCGGCGCGC 360  
60 ACCAGCGCC GCAACDAA GCGCGGGGCT TGTCTGACC AMCGCGGCTG GCGGAGCCA 420

GGAGAGCCC GCGGCGAGG GCGCAGTGG AGGTCCCGC CAGCCCCCG CCGCCTTCCC 480  
GCGCGCGGC GCTCCGAGG CAAGCGAGG CCGTGAGAG TCCACTAGT TCACAGGATA 540  
5 AAATCCACA CGAGACTGG GAGTCAGAA TGGCTAGCC CCAAGTGGT GTAGCTCTG 600  
TATTAATGC TTAGCTGCT GTGAATGCC CTGAATTTTA CCGTCAAGT TATTTCTCA 660  
GTTCACAGA ATCTATAGG GATGTTGAG AGGATTAAC TACTCTATCA GATATGTC 720  
780 AGGATTTTT GAATCATCT ACAGAGAGC CTGGAGTTT TGAAGCTGA ATTGAACAT 780  
TTGAGAGC CCGTAATGTT TGTGTTACA CAGATGATC TTTCAGAA CTTGTGAGC 840  
15 TCATCTATCA ACAGGCGCA TCTATCCCA ATTTCTCTTA TATGGAGCT CCGCTGTGTA 900  
ATTACTGTC CCACTATCTG ACAATTAGCC CACAGAGTG CAAGTCCGC CAATTGCTAC 960  
1020 TTCAAGATG TGGAGTGA TATGAAGTTA AGATCAAGC TCGAAGAGG GATGAAGTTA 1080  
CTCGAAGG ATTTCAACA TTGTACTCT TTTCGGAGA ACTTATCTT AACCTGGGA 1140  
TCAAGGAGC AATGAGAG GTTACAGAG CAGATATCT TCAAGTTGCT CTTGAGAT 1200  
25 TCGTGAATC CCGTTTCTT AATCTATG ATGACATTT AATTGTGCA GTAAATGTT 1260  
TAAAGTTGAC AGATCAGTT TTGGAGATG CTTCGAGGA AAGAGGAGG ATGATATG 1320  
AAGAAATAT TCAGAAAT GAAAGCTG TCTAGATGC AACTGCAAT AGAGATGTA 1380  
AACAATGCT CTTGAAGCT GTAGAACTC GTCAAGTAA CTGGGCGAG GTCCATGCA 1440  
CTTCACATA TAGAGAGCA ACACAGAAA ATGATCTTAA CTACTTTATG ATGAGCCAA 1500  
35 CATTATATC ATCTGATGTT GTTCTTTTCA CTCAGCTGA TCCAGATTAC CAGAGAAAT 1560  
ACCAAGATT ACTTGAAGA GAGGACTTTT TTCCAGATTA TGAAGAAAT GGAACAGAT 1620  
TATCCGCGGC TGGTATCCA TACTTGGATG ATATGATGA TCGATGGAC CAGAGATAG 1680  
AAGAGCTTA TGAAGTTT TTTTGGAT CAGAGCTTA GCGAAGCAG TAAAGTTAA 1740  
TTTCAGATA TCAGTTTAT AAGCAGTTT AGGTATGTTG ATTTAGGAGA ACACAAGGA 1800  
45 GCAAGAAAT GTGTACATC TATACCAAT TAGGATGTT GAGTATGTT ACTAATGAT 1860  
GCACTTTAA TTTTGTAAA CACTATCTGC CAATATAAC TTTATTCCT ATACCTTAA 1920  
ATGTATAT ATATATAATA GTTTATATG TACAGTTAAT TCTACTGTT TGGCTGCAAT 1980  
50 AAATCGATT TTGAATAAA TGAATGTTG AAATTTTGC TAGTTGTTA GATGCTTATC 2040  
CTTTAATTC TACTTTCTT GAGGGGAAA AGCTCTGTC TCGAATACA TATTAAGCA 2100  
55 AAATGTAGC ATCTTTTTT AGGTAGAGT ATTATAGCT YCATTTTACT TKGACATTA 2160  
GTGTCCCAAT GAATGAAT TCAATATGA ATCAATATCT TGAATATCT TAGCACTAA 2220  
GTTCTGGGA TATATCAAC ACTGATTTAC ATATGAGAT GCTATTTGNA TACCAAGGC

TTTTTAATG TGTGGGGGG GAAAAACCA ACTTGATGA ACTGCCACT AAAACACCA 2280  
GACTGACG GAGATTAT TCAATTGA GGAATGTC TTTTATTT TATTTTTC 2340  
5 AACGTCTA CTTCATAC TTAAAGCTA GACATTAT CTCTTATT TATGAGCTT 2400  
TGCACCTTT ATTTCGAC AAGGTGCA TGCATTTTT TACTAGAA TTTTACGA 2460  
TATTTCAT TCAATTTTC ACTGTCTA TATTGATAC AACTGAGG TCACTTTCT 2520  
10 AATCTTTA CTATATGTC GTACACCTC AACTCTATA AATATATT TTACTTATG 2580  
TCAATCTTT TTCCAGCTA CTAAAACTG TGTACAAAG GATTCCTGT AAATGCAAT 2640  
GTAAATAGT CTGTATATA CCACTGTT TACATTTGT ACATCTGT CTGTATATC 2700  
15 AGTTACCTT CTCACCTTC TCTTGTTTG TTCACTGTA ATTAAATTA GACTTGAAA 2760  
ATTAAGCTA AAAAAAAA AAAAAAAA AAAAACTCA G 2801

(2) INFORMATION FOR SEQ ID NO: 97:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1631 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ATGAGCCA AGACATAC TATCTTTC GCTTCTACTA TATTTAGG TGTCTCTCT 60  
35 AATTCTTC CATACATGT CATCTCTAC AGTATGTC CATGATGTA ACTGCTCTC 120  
GACTGCTTC TACTTCAGT TGTCTGCA GCAATCTCA AACAGGACA CAGAGGACA 180  
40 TGCATAGG GCTGCTGCG AACTGACT GTGACCCCG GATCTCTCT GACTCTCAT 240  
TCTTATTT TGGAGACA GAGAAAT GAAAACTG CAATCACA AGTTACAT 300  
AATCAGAT CTGAAATTA CACCTTAT CCTTGATGG GAGAGGCT TCACTCAGC 360  
45 CACACACA TACTCACA GAGAGGCTT GTTGATTTTC AACTTTACG CCACTTTTA 420  
AATTTCAC TCAAGATAT TCTGTGAT GTCTCATGT GTATACAT ACTGATGCG 480  
50 AACTCATCT GCTTACTTT ACCAGATT GCTGAGCTT GATTATGTC GTTTTGAGG 540  
GGAGCTGCA AAATCTGTA GCTTACACA GCTGCTTGG GTCTCATGT TTCTGCTCA 600  
ACCATAGG CTGTAGGCT GATGTGCA TGAATGCTC AAGACACAG AGTATGTC 660  
55 CAGAGGTA AAGATGTC TGTGATATC ATGAGACTT TATATGTC GATGCTGTC 720  
GCTGAGTG TCTCTCTCT TGTGAGCTC CTGTTGAG TGTCTATGT GCTGCTCTG 780  
60 AAGGCTCTT TGAATCAG TCTCTTTTT TATCTATGCG AAGACTGAGC ACTGAGATC 840

CTGATCCA AATCATGCG AACTTACA TTATGATG CTGATGTC GTTGAACT 900  
GTATTCAG AGTTAGAC AAATGAGTC CGAGATGTC AACTTACA TATTTTGT 960  
5 AAATGAGC CTCCGTAT CTCTGATG TTCTTTCC TGTGTGAC TTATGCTATA 1020  
GCTCTGTC TTGTCCTT ACTAGTGT ACTGAGAA TCAAACTT ACTCATGCG 1080  
10 GGAATTATC CATTTTACT GATGTGTC GTATGATG GATTTGTC CTTCAGATC 1140  
CGCAGTTA AAGGCTTTA TGAATATT AAAAAACA AGTACTGTT GGTTCACCA 1200  
CTGTGACT ACAGACCA ATGTGCAA CAGCTCAT CTCACACAC TCCAGATCA 1260  
15 TCCAGAT AAATGATG TGTACACAC TTGACTTCC CATTATGTC TCTTTTTG 1320  
TCACTCTC TCTTGACA TTTTCCAG TGAATCTCA GCTGTGTT TATGATTAAT 1380  
20 GTATTGACT TGTGTCTCA GCAATCAGG ACGAGGTC TAAATGTC CTGTCTCC 1440  
TGAATCTCT GACATFAC CTGTCTGTA TTTGTATG TGTAAATCA AGTCTCTGA 1500  
25 TACCTTAAA CATTGATTA AACAGATG GCAATCTCA TCTTAAACA AATGATAT 1560  
TATTTTATA AATCATGTC TCACTTAAA AAAAAAAA AAAAACTG AAGGAGCTC 1620  
GTATCCAAA T 1631

(2) INFORMATION FOR SEQ ID NO: 98:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 504 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

CGAGTGGG CAGATATG GGAAGGAC GAGCGCCCG GATGAGGTT GATATCTT 60  
45 CGAGAACT ACTGAGCG CAGCTGCA AAGATGAG GAAATGTC CTGTGATGTC 120  
TGCAGAGG ATGATAGG TCAACGAA AATAAAT CCCCCCTT GCTTCAATGT 180  
GAAAGCAC GTGAGATG TGGAGTGA TATATCAC TCAATGTA CACAGATAT 240  
50 CATCTCAT GTATCTGT TGCATGAT ACCAGAAC ACATATGA CAGTGTGTC 300  
AAGGTGTT GCACTGTGA CAGCATGTC CTGTCTTCC GAGGAGCTT TTATTTACG 360  
55 GAGCTCTG TTCAATCCA GCGATCTTA CCAAAAAAG CTTGTGATG AAAAAAGA 420  
AGTTGTGA TTTATATTA CTTTATGT TGAATAGAG TATTAACAT ATTCTGAT 480  
60 TCTTCAAAA AAAAAAAA AAAA 504



## (2) INFORMATION FOR SEQ ID NO: 99:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1416 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GGCAGAGGG AGGAGCCCT CTCGTTGGG TGACTCTGT GTGCCCTTTA GACAGGCTGG 60  
CTGCGCCGTT CCACAGGTTA CAGTAGGAC TTGAGTCTTT CTTTTCTGT TTGAGTTGG 120  
TGAGTAGTG ATAGGTTAC ATGGGCTTC AGGATGACC CTTGGACTG TCCCGAGTTC 180  
CTTAATCTC AGCTGGATC CTGAGCTGG GAGGCCCTG TGAAGGCCAG CTCGGANAA 240  
ACCTGGAGT TGAATCGGA CTTGTGGAG AACTCTGTC GAGGCCAGG TCCCTTGGA 300  
CACTGTACT TCTGGGCTG GAGGGAGAG GGGCTCGGC TTTCTTGAA ATGAACTG 360  
CTCTCAGCA GTTCAAGTAC TTGTTCTCA AACATTTCT AATTGATTG TAGGTTTCA 420  
TAAGCATTT TCTTTAAG CATTGAAAG GAAGATGCT CAGCAAGTC ATGTTTGT 480  
TCACTGGAT GGGCCGCT TCTACTGCT GGGGCTTCC CTTCTATG GACCTTTGT 540  
GCAAGGCCA CCAAGCAGC TCTTCCACC TTCTCCACT GAGCAACCA GGGCTTGGA 600  
ACCTTAATT GCTATACAG AGGATTTTT TTTGCTTAG TACTTTTAC ACTGTGCAA 660  
CGCTCTATT TTTTAAAG TTCTCTGCT TGTATTACA GGAAGTAGT GAGAAATAGT 720  
TTCTGAAGC AGTTTATTGT GAAGATCCC AAGGGAGGT TGGTAGAGA AAAATAGTAA 780  
GCTGGTTAG AACTGACGA GGGAAACAG CAGAGAGCA TTGGAGAGA ATTTGCCAAA 840  
GATCTACCT GAGATACGC CTGTCCAGT TCTTCACAC GTGATTAAC AGGCTTCAA 900  
AGTGTTTTC TCGTTTGAA AAAAAATTC CACAGCTTT TAAAGTGCA TTTAGAAATC 960  
CATGTGACT TAGAATGGA CTGCGGCCCT TGGCAACTGT CAGCTGTGCT AGAAGTTGG 1020  
ATGCTCTGG ATGCAATGT ATACTCACT CATTTGTT TCTTGATG CATTTTGTT 1080  
CTTTAGCAG ATCTGTCTT GTGGTGCTG TCTAAGAGT CGGACACTT GTTTTGTG 1140  
TTAGATTAG CTGGGAGCT GCATCAGCT TCTTTATNG CAATTAGGC AGCAACATC 1200  
TGTGTTCTT GTTGTGTC TAAATAGTG AGGGAGGGA GGAATGTAC CCCAAGATA 1260  
GGCCCTCCA TTGGCTTGG CCAAGCCAGA CACTTCACAT GTTTACATG GTTCTGTGA 1320  
ATTTAAAGT TTAATGTAT AAAGCGAAG TGTTCCTG AACTGTATA TTTTGTAAAT 1380  
AAATATATT CTACTTGAA AAAAAAAAA AAAAA 1416

## (2) INFORMATION FOR SEQ ID NO: 100:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2847 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GGCTAGGACA ATTTGCTGC TTACTCTATC TCTGCANAGA CTGGAGATT TGGCATACCA 60  
TTAATTACAA CCACCAATCA TATCCACAA AAGTACCCTA AAAGAAGCAC CAGTGGCCAC 120  
TCTCGAAAA ATTTAAGTAT CAGAAGATTA AAAGATTTT AGGATTTGGA AGCTGTATT 180  
GTCTTTCCCC AATTATCAAT GTTGATCTC CAATAGTAG CTTATATTA GCANTGACA 240  
GATCATGCTT TCTCCATATC TCATCATATG TTACTACTTT GGAATCAGTA TTTGGCANA 300  
TTCAAGCAAT TATCGATGG ATATNAATGG AAATATNAA ATATTTCCA ACCGTCTCA 360  
GTACTCTATC ATACTCTGT GATCTCTCA GGAAGCACT TTTGCTTTTA CTTAGAAAC 420  
GTTTCAGATT TCGTTATAG ACTCTGCTG TCTTCAGTAC CTGATTAAC TTTAACCAG 480  
GAAGCATTA ACAGATGCA GCAGCTTTTG CCCAGCTTC TAACTCTCG CCGCAGCAT 540  
TTATCATGT AGAGTAGG ATGCTTCTG CAGTGGCACT ACTTCCCT AGAGCTGGAG 600  
CATGCTGCTT GGCCTTAAGC CCCAGCATGA TGAAGCTTCC CTCCTGCCAG GTCAATAAA 660  
GTTAGAGAC TCAAAATGG GTCTTCCCTG GGTCAAGTG GCAAGGTTTG CTGAAACCCC 720  
TAAAGAGAG TCACCAAGG AGGAGGTAA TGAATGTTT CAGATCACT CAGATCTCA 780  
TAGCAATTC TGGTACTT TCAATGTTG AATTTCTGA TCGTAGAGG GACTTTGATT 840  
TGATATAT AAATCCAGA CAGTCCCAAG AAGTCTTGG AGTCTGGCT CTGACAGCC 900  
AAGAAAGGA ATACTGTGA TTAGGACCA ACTATAGCC AGGCCCTGAG CTGTCTCTTA 960  
GATATAAA CAGATGGGA GTGGAGAGT CATTTGCTT AAGTTATACA GCTAGGAAT 1020  
ACTCAAGCA AATCTGAAC GCAGCTCCC CTAACTCTGT GGAAGGCAC TTGTACCAC 1080  
ACACCTGCT CCACTAATA ACAGAGGAT AAAGACTT CAGGTTTTCC CACTGTGTC 1140  
TGACCATCC AATTTATGA TCTTCTTCA ATGCAATTT CACAGTATA GTTAGGCTC 1200  
AGAAATGCA TTGAGTAGC CTATTTCTC CCGTTAGCA GATCTTTAA GTACACATG 1260  
CTGACTGAG CCACCCCCA GAGTTAGGA GAACTTTTCC TTTTCAAGC CACTTCCAT 1320  
AAATAGGTG TTTCTGGCC TTCAAAGTA TAGAATTTG CAGCACTTG AAAGTCAG 1380

	GGTGTCTGTC TCTTACTCA ACTTATTTTG AAAATGTCTG CAGCTTCACT CCTGTAGAAA	1440
	AGGAAATCTT CATTTTTHG TAAACTTHAG CCGCAGTGTG CTCTGTGAG ATGTGGCAAT	1500
5	TCAAGTSCA GTGAATCTG CTCTCTTACT GATTCCTGAT TTTAGTGTGT GTGTGGGGGG	1560
	AGTGTGTACC TATTTATAAA GAGAGAGTGT GATATGTGTG TATATGTATA TACATACATA	1620
10	CAATGTCCAA CACACACACA CAAATTTTGA GAGCTHAGCA AATCTGAAA CAGCCCTTTC	1680
	ATTATCTTGC GTTACTATTC AAAATTTTCT GTCAAGCCCTA ATTACAGTGT TCACATATTA	1740
	GTGTGGGGCT AGGTACTTGC TTACAGAGAG AGCAATTTCC TACCAAGTGT CATTTACTTC	1800
15	TAAAGCATGT AGTCAGATGT AGACCCCTGA AGGAAATTTG ACTCCAGCCC TCTCTCAGAA	1860
	TCTCTATATA GATGGGAAAC TTGATATGCC AGCCATTTTG GTGACTGTAG AGCTHAGTA	1920
20	CTCCAGTTAG ACTHAGAGGC ACATATTCAG ATTCTATGAC CTGTATGTGT TGGCAGGGTC	1980
	TAGAGATGCC TCTCTCCCCA AGTGAATAAT ATTCTTTTGC CATTCCTGAA ATTCCACATT	2040
	CATHTATATG CTGTGCATTA CATGCTTCTC AATGAGAAA TTAACTGCAT GTTTACTGTG	2100
25	TGCTGTATAC ATTCAGATTTT TATGTTTTAA AAAATCTCAT TATGTATTTA GTTCAGCCCA	2160
	GCTHAGAG AAAAGAGAG CCGATATGGG AATCTTCACT CTCAATTTTA TTGCCCTTAT	2220
30	CGAGAGTGC TTATATAGGC CCGTACCTTG TCCATTTCCA GAATCAGATA GATCTCAGCA	2280
	GTTCCTTGAT CTGAGAGGCT GCGTGTATAG ATTAAGATAT ATTAATTTGA ACTGAGAAC	2340
	TACAGAAAC AGCAGAACCA GCGCCAGAGC AGGAAATTTA AAATATATGT AGACATTTAT	2400
35	GATATATTTG GTCCAAAAA AGCCAGGGGS CCGATATCTG GCTTATATGT ACTTTGAGGG	2460
	GACACAGAT TAAATCTTAA TGAAGTTTAA ACCTCAGCAG TCTTTCTACA GTACAGGCC	2520
40	ACACTGTCTG AATGGGAGAA AGCAATGAAAT CCAATTTTCT CTGCCATTTT TCCCTGTGAC	2580
	AGTCACATTC CCGCTTTTAG AATCTTCCC TTCCAGCCCTT TACATTTAAC AAGGAGAAC	2640
	TGAATCTTTC AAGGAATTTA CAGCTTTTGG TTAATGTTTC AGTATATCAT TTTCATATCTG	2700
45	TAAATTTATTT TGTAAAGAG ATTTTACTGT ATCCAGAGAT GTTCCAGATT GTGTCCCTGT	2760
	TGCATTTGSA AATCAATTAA CTATTTACTG AATTCGAAA AAAAAAAAAA AAAAAAAAAA	2820
50	NNAAAAAAAA GAGGGGGGCC CGTACCC	2847

(2) INFORMATION FOR SEQ ID NO: 101:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1394 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

	GAGATGTCTG GAGAGAGTAA AATATCTAG AGGAGAGAT TGAATAGAG CCAAGGGGAA	60
5	CCCCAGAAA AAGATATGAA GCTAACTCAT CTCTTTTACA AGGGGTGGCC ATATCTTACT	120
	GTTCAGAGT ACTCAGATTA TATTTAATGT TGAATGTGAA ATTATGATTA CGAGAGGAAA	180
10	GAAATTTT ACTGTGTC TTATTTACT TGTGAAAAG CTGTGGGACA AAATATATG	240
	AATAGAGAG GGCATCTTCT TTGTATTTTC TGTCTTCAAT GCATHTATTT TTAATTTACC	300
	ATAATTTCCA AAGATTTTGT CATTGAGCTC TCTGTATTTT TTCTTCAATC CAGCCCTTTC	360
15	CTTTTTTGG AAGGGGGGTTA TATATAGAG TTCAATGAGG AAGTCAATGT AGCCTGAAAT	420
	AAAGGGGCAA GATAGGGGAG TTAACTHAGG AGCAATTTAT TTCTTTAGG CCTTTCTHAG	480
20	AAAGAAATGT GGTGTGAGAT GGTGTGAAAT TCCCATGATG TTGAGAGGCA CTTAGTGGGG	540
	TTGAAATATG AATATATTTT TCCCATTTGG GAATAGAGAA TTTCCTTTAG AAGGTGTGAA	600
25	AATGCTTTTG CCGAGTGTCC CATTTTTHG CATCTTTTCC TTCTTATTTG CTTCAGATCA	660
	GATGTGTGTC TATACAAAC TTCCCATGAG TTCTCTGCAA TATTTCCCAT TTGTAAATGA	720
30	TCACTTCTCT TTCTTAAAC CTTTTCTCT TCAATCAT AGAGATTTTG CAAGGTATAG	780
	ATCTATCATG CAATTTCCC TTCTTCACT GTGTCTTCTG CAATCTATC TCAATAGAAA	840
	TTCTGGGTGT GACAGAGTAA GCTAAAGTTT GGTGTGTGGA CTGAGATTTG GCAATTTAGC	900
35	ATCTATGAAA TTCCAGTCTC CTTCACAAA GCGCATCTT GCTHAGTGC ACAGGGGAAA	960
	CGTACGTCCC TTGTGGGGCTC AATTTTCCCT CCGCTCATGT AAATGAGAAA AATTACTATT	1020
40	TTTCTTTTGT GTTACCATT GCTGTGACAA AAGTGTATCT ATTATGTGTG TATTTGGGTA	1080
	TGTGTGAAA ACTGTAGAG CTCACTGCTT ATTAAGAGAA ATTAAGAGAA AAGTGTAGAA	1140
45	GAGAGACAAA AGAGATTAAT AATTTGTATA GATTCACAAA TCCCACTTT TCTCTCTGCA	1200
	GTCCCTGTCT CTATGTTTTC TGTTTTGTGT AGTCTTTTGT GCAACAGCCC ATTAATCTTT	1260
	GCATGTCTGC ATCCGTGAAA GGGGGTAAAT GGTCTCACA GTTGTGTGCA TTGTTTTTTT	1320
	GCATCTTCTG TTAATATAAA AAAAAAAAAA AATGTTTAGG TTTTATCTTA AAAAAAAAAA	1380
50	AAAAAAAAA ACCC	1394

(2) INFORMATION FOR SEQ ID NO: 102:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 794 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(D) TOPOLOGY: linear

(xi.) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

5 GGRRCGAGGC GAGTAAAGG GACTTGAGC AGCCAGTTGC CGGATATTC TATTTCCTC 60  
CCCTCTCC CCGCCGTAT CTCTTTTAC CTTTCTCCA CCTTGCCTG CTTACATCG 120  
CGAGCTCG CGCGCACTC AGTCCCATC CATCTCTCG TGTCTCTCG GAGCCGAGCC 180  
GTCCGCCCC GCGCGCGCG GAGCGCAGG AGCTGCCCC GCTCTGGGA GATGTGAGA AGACTGGTT 240  
CAGCTCTCC TGTGCGTGC AGATCTGAT TTTCTGGAGA GATGTGAGA AGACTGGTT 300  
TGTCTTTGA CAGCTGATC ATGCTGCTT CCTTGCAGC TTCTAGTGC ATCATGTG 360  
GTTCTTAC TCATCTCG TCTCTCTCT CTCACATCA CTTCAGGAT CTACAGTCC 420  
GTCACTAAG CTGTWAGAA TTCAGAAA GGCATCCAA TCCAAAGCT ACCTGAGCT 480  
AGACTTACT CTGTCTCAG AGCTTTTCA TATTACATG ATGCTGCGA TGTGTGAGT 540  
CAACGCGCC CTGAAGCTA TTATCTGCT CTTCTGCTA GAGATCTCG TTGACTCTTT 600  
GAAGCTGCT GTCTTCATGT GCTGTATAC CTATGTTGT GCTGTTTTA ACCGATCAC 660  
CCTCTAATT CTTCCTGAC TCTCTATTT CAGTGTCCG ATTGTCTATG AGAATACAA 720  
GACCGAGTT GATCACTATG TTGCAATCG CCGAGTACG ACCAGTCAA TTCTTGAAA 780  
GATCCAGC AAAA 794

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1544 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi.) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

45 TTTCCTGCT AGTCTGACC AAGAGTTGT TTGGGCAATT GCTGTGTGG CCATTTCTCG 60  
ACGAGAGGG TCTTCTTCT CTTTCCCCA GCGAGCGAG TGTCTTGGG CCGAGCTTTC 120  
CTGGCTGGA AGAGTATAC CTTTCTCTG GCGCTTAGA TACGAAATG ACCATATGT 180  
GCGCAGCTG CCTTCATGC TGGCGCCAG CCGAGTCTG CACTGCGCT GATCAGCTTC 240  
TTTAGGCTT AGCATCTCC TGTAGTAG GATGAACTT GCGAGCTTC AGTTCGTTTC 300  
AGCTATGAC ATCTGTGGG TGAAGGTTACA CTCAGTCTC CTCCCAACT CCGAGAGCT 360  
TTAAGAGTG TCCCTTTGC GCGCTTCTGA GCGAGAGAC TGAAGTGA CCTGTGAGA 420  
CTCCACAGG GAGGAGCGG CTGGCTCAG GATGTGGGA CCGAGACTG CCGAGGCTTT 480

5 CAAGAGCCT GTGTGGGGC CCCAGGATC CTTAGCTGAA GCGGGGAGAC TCACCTCTCA 540  
TCTGAGAAA TTCTAGCCTT TCCCTCTAG GAGCCAGGCT TGAAGGTTAG GCCCAGAGC 600  
TCCCTTAGG CCTTGGTGG GCACTCTGG CCGCTGGGCT AGGAGGGGAG ACTCAGGCCC 660  
ACACTGGGT ATTTCATAT TTGAGAGAA CACACACTCA GCGGCGACTC ACTGATTCCT 720  
10 ACACATGCC AGATTTTAC ACATGTGAC AGGGGCGACC AAGTCCCTG TCACCTTTGT 780  
GACTAGATC CTATTTCTC TATTTTCCT TGGTGGCTG GGTCTGTCTC AGCTGGGGA 840  
GTGTGGATA TGTTAGTTC TGTGACTG TTTTGTGGG GTGGCACCTG GTTCTCCGAT 900  
15 GCGTGGGCTG GTGTAGGCC CAGGACTGTA GTCTGGGAG CAGTAAAGT CAGCTCTGTG 960  
TATGTAGTA TCTATGCTT TGTCTGTCT TTAGTATCA ATCTTTTCT ACATCAGCC 1020  
TTGTTTGTG TTATGCTAG TCTATCTGG CTTGTTATT TCTTGGCGG GAGGAGGGG 1080  
TTTCTAATC TCTTCCAGC CCACTTATT ACACCCCAAC CTGCTGGGA CTTACTGCTC 1140  
GGAGGCGAG AGACAGGGG CACAGAGAG TGGTTCCTG CCGCTGTCTT GGGGTTGGG 1200  
25 GGAAGCTGG GGCACATGT GCGCTTCCCT TGTGAGCAG TCCAGTGGC AGGCTTTGA 1260  
GACTTTCCA CATGATAAA GAAAGGGAG GTACAGAGT TCCAAATCC TTTTATTTT 1320  
GCTGTTGCT ATCTTAAT GTTTATATA TATCTAGCA TGTATCTATC AACCCAGAA 1380  
ATTTCAGAT CTTCTCAAC AATATGAGC TTTTAGCATG TTTATATTC TTCACTCTC 1440  
35 TTGTTTCCA GTTTTGAG GGAAGAAAG TCTGAAATTA TAGATACAG TTATTTATTA 1500  
ATTGTCTCT GCATAAAAA AAAAAAAA AACGCGGGG GGGG 1544

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 871 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi.) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

50 ACCCAGCGT CGGACTGTG CACCGGGGG CTTGGAGTG AGTTACAGA TTCAGCCCAT 60  
TTGGCCCCG CCGCTCTGTT CTGGGANTCC GGTGTCTGG GATTGAGTTC CCGTTCTTA 120  
55 AGTGGGTGG CTGTCCACCC GCGGGGCTG GAGTGAAGTA CAGATTTCAG CCGATTTCG 180  
CCGAGGCTT CTGTCTCGG AATCGGCTG CTGCGATTT AGGTTCCCGT TCTTAACGA 240  
CTCCAGATG GAGGAGGGG GGAAGCTAG AGGCTGATTT AGATGTGCTC ATCTACTGT 300

CTTGTCTAGGT GCGTGGGCGA TGCATATGTC GGTACAGCTC GTCTCAGGCT TTCTCTCTTT 360  
TCCGAGAGCT TCCCGGACAT ACCTTCCGAC TACGTCCAGAG CAAACTCTTC CCGTCTTACT 420  
TCCAGATCTC CATTGGCTGT GCGTTCACTA ACCTTCGATC CTTCGCTTCA CAGCATGCTT 480  
GGGCTCAGCT CAGATCTGTC GAGGCGAGCC AGCTTACCT GCTATCTGTC AGCTTACCC 540  
TGCCTCACTGT CAGCGCCGCC TGCCTCGAAC CCGGACAGAC AGCTGCGATG TGGGCGCTCC 600  
AAACCGCTGG AGAAGAGAGC AGCGCTGGGT GGGAGAGTAC CAGGCGAGCA ACAGGTTCCC 660  
GATCTTAAAC GCGAGTGGCG AGAAGAGAC CCGAATTTCA GTCTCTTCCG CCGAGATTTTC 720  
TTCCGCTAAC ATGGGCTGTC CTCTCTTTCC AATCTGGGCT GCGTCCGTAG CATTGGGCTC 780  
TCTCTCCGTC GCGTTCGCTT GGAATTTAGG AGCTCTTACG ATGGGCGCTG CAGCTTAAGA 840  
AATGCTCTCT CAGAAAAAAA AAAAAAAAAA A 871

## (2) INFORMATION FOR SEQ ID NO: 105:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

GCGAGAGATT ATACATAGCC ATTGATACCT TTGTTTAAAT GCGTCAGTAC TTTTTCAGAT 60  
TTAGAACAAA ACATGCCAAC CCGTAGAGCT TCTTCCGATG AATTTTTCGA TCTGCTCCAA 120  
AACTGCTTTC AGTTACTCAG AACTTCAGCC TCCGATTTCA CTGAGAGCAT TCGTTTTCGA 180  
AGATACAGAA ATGGGTTTCA CATTAACTT GCGAAGATTT GAGAGAGCTG TAAATGTTTC 240  
TTTTTAATTA GAATGAGCC CCGCTTTGCG GACTAAATTT GTGCTATTCG CAGAGAGCAG 300  
TCTAAATTTT ATTTTTTTAA AAAAGAGAC TCCCGCATTA TTTTGGTGGG GTTGTTTTTT 360  
AATTTTAAAT TTGAAAAAAT TTTTGGGGGT TTTTGGGAGC ATGC 404

## (2) INFORMATION FOR SEQ ID NO: 106:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

GTGACAGAG TGCAGCCGCC GGGGAGAGAG TCTGAAAGAG CAGAGCTTCA GAGAGAGAG 60  
GCGCTTACGA GAGAGAGAG GAGAGCCCTG GAGAGAGAG GCTGGAGATC CATTAGCCAG 120  
GATGGAGGCT GTTGTAACT TTACAGAGA GGTGATGAGG CAGCGAGATC CCGGATTTCA 180  
GGGCTACGCT CTGATGGGAT CCGCTTCTCT AATGACCTTC ATTCTTTCA CTTACGTTGA 240  
CTTGTGTTTC TGAATTTGGC CTGCAATCAT GGTATATTCG AAGCGCTTTC AGCTCCGTCG 300  
CTTCAATGAT TTCTACAACT TTCTACATGTT GCGACATCTC CTCTACATTC TCTATGAGTT 360  
CTTATGTTG GCGTGGCTGA GACATTTTAC CTGGGAGCTT GACCTGTGTC ACTATTTCAA 420  
CAGCGCTGAG GACATTTTGA TGGTTGGGAT GCGCTGAGTC TTCTCTTCTT CCGAATTTCA 480  
TGAAGTTAG GACATTTTGA TCTTTATCT CCGAAAAAAA GAGGAGAGAG TGAAGTTCTT 540  
ACATGCTTTC CAGTACTCTG TCGTTCCCTG GAGCTGCTGTC TGGGGGGTAA AATTTGCCCC 600  
GGAGAGAAAG GCGTCTTTC ATGCAATGAT AACTCTTCC GTGGATTTCA TAAATTTACT 660  
GTACTTACGA TTATCTGCTT TTGGGCTGTT GGCACAGCC TACCTTTGAT GAAAAAAGCA 720  
CATTAGAGCC ATTCTACCTA TCCATTTTGT CCGTGTCTCA CTGCAATCT CCGAATTTCA 780  
CTTATGTTTC AGCTTAACT AGCGTACCC AGTCATTTAT CAGCTGATCT GATGTATTCG 840  
CAGCATCTTC TTCAATGCTT TCTCAGATT CTGTATTCAC TCTTATTTCA AAGGCGAGAG 900  
GCTGGCCCCG GCACTTCAGC AAAATGAGC TCCAGGTTAT GCGAAGTTCA AAGGCGAGTC 960  
AGAAGCATTC CTATGATTCG CCGGACCTTA AGTCCCTGAG GACTGTGACT TTAGGGCATTC 1020  
TCCGTCAATG CCGCTTCCAC CTACACTGT GAGCAGAGCT TATGTGATTA GACTGTACGA 1080  
GGGAGCTGAG CTTCCCGTCC CAGAGAGTC TCTACAGGGA CAGAGGCTTT GATTCTTCAC 1140  
CGATTTCCCG GGGGAGATTC CAGGATTCG GGTCTATTCG TTCTGCGAC TCCAGAGACTG 1200  
GGGCTTAAAA GCGCTTTTCA GTTATTTCCC CTTCCCTGAC TTAAAACTTC GAGAGGAGAG 1260  
ACTCAGAGCT GCGGCGAGAA AGGATCTGAT GCGCTTTTTC GTACAGAGAA AAGATGAGC 1320  
AATATTTTCA CTGTGAGACC AGTCTGACTC CTGCAGCCCA CAGCAGTAAAG CAGTTAGCTTC 1380  
TGGGCGAAAG GTCAAGGTCG GCGGGGAGCT GGGATTTACG CCGTGGAGAG CTGCTTACTC 1440  
AACTGTGTC TTAAATTTAA GTACAGAG GAGACAAAA AAAAAAAAAA AAAAACTTCA 1500  
GGGCGGCGCG TACCGAATTC GCGGTTATTA TCTTAAAGAA TC 1542

## (2) INFORMATION FOR SEQ ID NO: 107:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2127 base pairs  
(B) TYPE: nucleic acid

## (2) INFORMATION FOR SEQ ID NO: 107:

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

5 GGTAGCTCAN TCGAGTCGAA TACTCTTACT GGAACGAAAG CCTTTATCA AGATATATTA 60  
ACTCTTCCTT TTCTTTTTCG GAGAGTGTCT TTGTTTCTGA TCGGACCAAT TCACTGCAGC 120  
10 AAGCAACACA GTATTTCTTAG CAGAAGATCG GGACTTCGAG CCATGTTTCG GAGGGCCAGT 180  
ACATTTATCT GGACTCTGGA GTGTGAGGAA TATGAGACTCC ACTCTTCACT ATATTCACAR 240  
CGATTCAGAC TTGAGCAACA ATAGCGATTT TACGCTTGAT GAGGAAAGGA GACTTAAGT 300  
15 ACAAAGATGT GTAGCTCAGG CGTTCTTAGA TCAATATTTA TCTATGACTG ACCCTTCTCG 360  
TCGACAGAGG GTTGACACTG AAATTCCTAA GCACTGTGCA TATAGGCTCC CTGTGTGCGC 420  
20 CTTGACACTC GGAAGACAGA ATTGCACTG CCTGAGAGAG ACCATATGGA CTTTGGCCTC 480  
AGACATGCGA TCGAAAGATTC GAGGAACTC TAGCATCTTC CATCCAGCGG CTTCCAGTTA 540  
TTCTTGAGGA TCATTTGACA GCTGAGACT TGGTTTCAAT TTTTAAATGA TTTTAAAG 600  
25 ACCTGATGTA AGTCAGATA GGTGTTCTTA AACACTTCCA TGAATTTCTG AGCTTCTTC 660  
ATATTGACA AAGAAGAGAA TATCTTTATC AACTTCAGGA GTTTTTCGTC ACAGTATA 720  
30 GTAGAAATTT GCGGTTTCCA GCTGACTGCG CTGACAGCT GATTTTACTT CTAGAGTTAT 780  
ATAGTCCCGG AGATGTTTAT GACTATTATC GTCCATTCGC TCTGATCTCG TGTGCAACA 840  
ANGTTTCTTC TGTTCGTTGG ATTCTTACA AGTTGCTCAG CGAGATGCTG AAGAAGCTGC 900  
35 ACCGGGCAAC ACCACACAG TTGAGAGTGG ACCTCATCAA TGAAGTTTGT GAGAACTTGG 960  
CGAGATGTCG CAAGTGTCTT GGTGCGGAG CCTTTGTCTT TGTCTGCCAG ACTGTCAATG 1020  
40 AGGATGACTG CTTTCCCATG GACCAATTTG CTGTGCATCT CATGCCCAT CTGCTAACCT 1080  
TAGCAATGTA CAGGTTTCTT AACGTGCGAG TGTGCTGTCC AAAGACATTA AGACAACTC 1140  
45 TACTAGAGAA AGACTATTTT TTGCGCTCTG CCACGTGCCA CAGGAGGCT GTGAGAGAGA 1200  
CCATCATGCG TCTTCAGATG GACCTGACA GCGATGTCAA GTATTTTCCA AGCATCCAC 1260  
CTGCCAGTAC CAATAATCTC GAAGATGCCA TGAAGCAAGC GTCTCTAAC TACTAGAGG 1320  
50 CTTGAACTTC GGTGTCTTTC CTGCTTCAAT TGAAGCGAG GTTCAGTGGG CATTCGCCAC 1380  
CGATGTGACC TGGATAGCT TTGCGGGGAG GAGAGACTT CTCTCTCTCC GGACTTCATT 1440  
GZAGGTGAA GTTGCTTACA CCCAATACA GGAATTTCAA GAGTCAGAG AAGATACAT 1500  
55 AAACACTATT ATCTATCTT GACTTTAAG KRWNRHMM KCTGAGHSA TTHATFTSH 1560  
CWRNRGSH WYHAWSCYK SKECTYVCC KSRSTGRMG MHRCTCTACA ATTYGRLAK 1620  
60 CNYTKSCT RRMGGAHKS GZCAGAGCC AGAGACTTCC ATTGCTTTCT CTGCTTTT 1680

TTTAACAATC GACAAATGAA ATTCTTACAG CCTGAAGCCA GAGCTGTGCC CAGATGTGAA 1740  
5 AGAGCTTTC AGTATCAGCC CTNACTCTTC TCTCCAGGA AGGACTTCTT GGGCTCTGTG 1800  
GCCAGCTGTC CAGCCGAGCC CTGTCTGTGA ATCTTTTGTG AGTGTGCA ATGCGAAAGG 1860  
AGGGTTTTT ACATCTCTTA AGGAGCTGCA TCCCAACACA AGTAGAGTTC ACTTAACCTC 1920  
10 TTAAGCGCAG CATATGCTG TACACATTTA CAGAAATGTT CCTGAGTCTC TGTCTCTGAT 1980  
TTTTTCATGC TGGTCATGAC CTGAGGGA TTTTATAGAC GTATATGTTA TGTCTGTGT 2040  
TTTTTACTTG ATCATGATCA GCTCTGAGGT GCATCTCTTT CACATCACTGT ACATCTCTT 2100  
15 GACCACTCTT GGGAGTCTCG CAGTCTTTAA TCACTCTGTT TAAACTGTTC TGGCACAAGT 2160  
TCTCTGTGCC AAATTAATTT TATTAATAG ATCTATAGAG AGAGATATAT ACACTTTTGA 2220  
20 TTGTTTTCTA GATCTCTACC AATTAATGCA ATTTGTGACC TGTNAAAAAA AAAAAA 2280  
ACTGAGGG GCGCCGGTAC CCAATGCCC GATATGATCT AATCATC 2327

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1062 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

GGCGCCGAG GCGCAACAGC GTTCTGTCA GCTCTGGTTC CAACCGCACT AGCGAATATC 60  
TTCTCATGC TCATCATGCT CTTCCTCATC CGGATCTGG TCCAGGTCCC TCTCCCCCCC 120  
ACACAGAGG TGGCGAGGT CCAGCTGTAG TTCTCTGGA CTTTCTGAA GATCTCTTTC 180  
CTTCTTCTG TCATCATCTT CTTCTTCTTC TTCTCATGCC TCATCATCCA GTTCTGAG 240  
40 CCGCTCAGA ATCCCATCC CCGCGCCGGA GGAAGTACA GGAAGCGGG GTACAGCTCT 300  
TATCTTTCAC ATGACCATTA CCAAGGCCA AGAGCTCTAC AAAGAGAGCG TCGATATGAA 360  
GAAGAGGG TGGCTCTCAT TCGAAGATA CTTGCGGCA TCACTGTATC AGAGCTGAA 420  
50 CAGAGTTCT CCGTTTTTGG AGAGATTGAG GATGCAACA TCACTTCTG TGTCCAAGG 480  
GACACTGAC GCTTGTGAC TTATGCTAT CTTGAGAGG CATTTGAGC CATTTGAGT 540  
55 GCGCAAGC TCGCGAGGC AGATGAGAG CCTTTGATC TCTCTTTGG GGGCGGAGG 600  
SMTNCTGA AGAGAGCTA TTCTGATCTT GACTTCAACC GGAAGACTT TGAACGACA 660  
CCTGTAAAGA GCAATTTGA TTCTCTGAC TTGTGACAT TGTGAACA GCGCCAGAG 720



(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

5 AGCATGAGCG CGATGGCGCT GGTGGCGAGT AGCTCTCTGG GCTTGCTGCA AATCATGGCT 60  
GCTTTTGGCG GCACTCTGGT GGTGGCTTAC TACGTATTGG GCATCATTTGG GATCAGCTTG 120  
TTTAGAGGGG TCATTTGGGC TCTTCTGSEA AATGAGAGCC TGGCCCTCTG CAAATGGCTG 180  
GGCCCTCTGG GCACTCTGCA GCACTCTGAG TACTTGCGCA ACACTTCTGA TCACTTTTGG 240  
GCTGCGCTGG TCACTCTGTS GAATCTGATS GTGGTGAACA ACTGGAGGT GTTCTGGAT 300  
GCATATCGGC GCTACTGAGG CCGTGGCTGC AAGATCTAAT TTGTATTGTS GTGGCTGTTG 360  
TGTCTCTGCA TCTGGGTGAA CTTGTTCTGS GCGCTGATTC TGGAGAGCTT CTTTCAAGAG 420  
TGGAGACCCC GCAAGCCACT GCAGCCCTTT GCTGGAGACC CAGAGGCCAC CTACAGATG 480  
ACTGTGAGCG TCTGTGTCAG GCAATATTCTG GAGGAGAGCG GGGAGAGATGA GCTTCAAGAG 540  
AGCTGAGCCC AGCAACCGCA CTTGTGGCTG TCGAGGTGAC GTCCCGGCTG CCAATCCAGC 600  
AGCGCGCGCA GGAAGAGAG GCTGGCTTAA CACAGGTGCC CATCTGSEA GAGCGCGCCA 660  
TCTGTGGCC AGCAGAGCAG GAAGAGAGCT TTCTCTGAC GAGCCACTTAA GCTGGGACA 720  
GGAACCAAGT CTTTTCCTGG TGGCCGACAA ACCATCTACA GAACAGTGC TGGTCTTTCA 780  
GGAGGCGGCC GTGGCTCTCG CTTTCTTTTA TAGCTGCTTC AGTGAGANTT CCGTCTGCA 840  
CTCCAGAGG ACTTTTCAGA CAATAATGCA AGAAGCAGCG GCTTCCCTGAG TCCCTCTGAG 900  
CTTTCGTTGT CCGTTTCTGTC CCGCGAGCC TTGGGAGCCA CAGCGCTGAC CAGGGGCTGC 960  
ACAGGTTAAC GGTGAGTCTG TCTCATCTAT TCACAGCTGG GAATGATCT AATACTCTCG 1020  
ATTTTAGCCC AGCACACAG GGTACGTTCC AGTTTCTTC TCTTTCCATA GCTGTAGGCG 1080  
CCTTCTGCGG AATGGTTCTC ATCTCTCTTA ATCTATTATT GGGTCAATTT TCTGTGATGT 1140  
CCCCAGCTTC CCATCACTGC CACCCACTCC CCACAGAGAT GCGCTGCTCA TCGAGCTGGG 1200  
GCTTTGACT CCACACTGTS TACCCCTCTT GTGTGGAGCG CCGTCTGCCA AATCTTTCAG 1260  
CAACAGCTT TCCAAATGSA AGTTGTACT GTACGGCTTT TACATCAGC AATAGCAAAA 1320  
TCTACATGCT GCTGAGGCTC GTGGCTCAT AGATGCAAT AATATGTAA GTACATTAHA 1380  
ACAGCATATG AAGAAAGCTA ATGCTTTAAT CTCAAATATG ATGTCTACAT AGAAGAGCCA 1440  
AATATTTTAA GAATATATAG AATTCACCA GCACTTTGGG AGCCCGAGGC GGGTGGATCA 1500  
55 TCAAGTCCAG AGATCCAGAC CATCTGCTT AAGAGGCTGA AACCCCTGCT CTACTTAAAA 1560  
TACAAAAAAT TGGCGGCGCG CATGGCGGCG GCGCTGTGCT CCGAGTACT GGGGAGGCTG 1620  
AGCCAGAGCA ATGGGCTGAA CCGCGGAGCG GAGCTTTGCA GTGAGCGGAG ATTGGCCAC 1680

TCCAGTCCGC AGTCCAGCCT GGGCGACAGA GCGAGACTCC GTCTCAAAA AAAAAAAAAA 1740

AAAAAAAAA A 1751

(2) INFORMATION FOR SEQ ID NO: 111:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1117 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AAATGTTGCG TCGTACGATT TGGTTAATTT CTTATTTATG AGTCTCTTGG AGAGCAATGT 60  
CCATTAACCTA ATCCCAACCA ACNTGTCTTT TTTRATGTTG TAGTGACAGAG CAGAGATTTT 120  
CAAGAGCCT TCGTAATATC TGTAAAGAGG CAGCTACAGC AGGCATCATT GCGTGGTGT 180  
ATGGGGGAT ACCAGCTTTT ATCTCTCTTA AACACATATA CATGAGCAG AGCCAGGCG 240  
AAATTTATCA TAAACGGTTT GATCTGTGCG AATCTGACA TGTGTCTGCC ACACAGGGCT 300  
TCATCTGTTA TGGCTGGCGC TGGGGTTGSA GAATCTGACT GTTGTGACT ATATTCATCA 360  
CACTGACAC TAGTCTGAT GTATACCGAA ATAAAGATGC CTTAAGCCAT TTGTATATTG 420  
CAGAGCTGT CAGCGGAGT CTTTTTATGSA TAAACCTAGG CCTGCTGCG CTTGTGCTG 480  
GTGGCATAT TCGAGGCTTG CTGGGACTC CTGTAGAGAG CCGTCTGATG GCATTTTCAGA 540  
AGTACTCTGG TCGAGCTGTT CAGAAAGAA AACAGAGSA TCGAAGGCA CTCATGAGC 600  
TAAACTGCA AGAGTGGAAA GCGAGACTAC AAGTTACTGA GCACTGCTCT GAGAAATTTG 660  
AAGTTAGTTT ACAGGAGAT GAACCTGAGA ATGATCTTAA GAATTTGAA GCACTGCTTAA 720  
ACCTTCTTAC AAACCTTCA GTATATGATA AACAGAGAA GGAAGTGAAG TCTCTGAAAC 780  
TTGAAACTCA CTCGAGAGCT GAGGGAGCT GGCATGCTCG ATGATGCTCA ACAGACAGGC 840  
CACTCTTGG TCGAGCTGCT GACAAATTTA AGTCTCTGTA CCGTGTGCG CAGTGGCTTG 900  
CTCTGTCTT TTCTTTTCT TTTTACTTAA GAATGGGGCT GTTGTACTCT CACTTACTT 960  
ATCTTAAT TTAAATACAT ACTTATGTTT GTATTATCT ATCATATAT GCAATACATCA 1020  
ATATATCCAC CCACCTAGT TTTAAGCAGT AATTAACCA TTTCGCAAA GATTAAAGTT 1080  
GAATTTACA GTTAAAAA AAAAAAAAAA 1117

(2) INFORMATION FOR SEQ ID NO: 112:

60

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1313 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

55  
50  
45  
40  
35  
30  
25  
20  
15  
10  
5

GGCAGAGCTT TTCTTATATT TTATGAAAT TTAAAGTGC TTGCAATA TTATCTTG  
TTGAGACTA CCAACATAC TACATGTGA AGCTGCTCA CAGAGAAAT ATTGCTTTA  
ATGTGAATA ATTTCACA ATGTGCTTA CTTATATTA GTATTAATAT TGTAAATAT  
TGTATTAAT AGTTGTAAC CCTTTCTAT TTATATAAA GTTAATGCA GTTACTTTT  
TTTGAATAA TGCATGCAAT CTCTTGAAC ATGAATGCG GCTCTCAAT ATGCTGCAAT  
TACTAGAAA AGAAGAAAG GCAAGCTTC TTATATATCT TGAGCCAAAT TGTAAAGCT  
TGTGCTTCAG AACGAAATCT GCAATTCGA AACAGCTGT TTAAATATGAG AAGGCAACA  
TTAAAGAAC TTACAGCAAT GCTCCATAG TTACATATGA ATTACTTGA TTGAGCTTC  
GCTATATTTA GCGAAAGCT ACTTCCATG CTCACAGCA TGAAGAGCT GAAATATATA  
AAGTTTACA GTTCCAGTATC GAAAGAGAC TGTGTGTGTC TTAAAGCTGA AGTTACATA  
TGTATTTT GTTACTGTCT TCTCTCAGCA GTTCAATATC TTTTGCAGG TTCTTGGTT  
TGCAGAGCAAT TTATGACAAA GCGAAGAAAG ATTATATCAG CAGCTAAA GAGTGACAA  
TTTATGATCT TTAGAGAGCT AGTTTGGCC AACTTAAGAT TTATGCTTGA TTTTACATA  
GTATTTGACA CTCATCCAAA ATATATGTGA AACATCTAGA TTTATATATT TATTTGCGC  
CTTTGTTTAA AACTGAAAT TTGTGAATAT GCTGTGCACT GCTCTCCAG CTAAGAAATA  
TTTTTGTCAA ATGGAAGCAAT GATTTATGCT GTGATATGAC CAAAGAGAC CAGCAATTT  
ATTCAAAAAC AATGTGTCA TCAAAATAT TGTCTAGATT GTTCAATATC ATGTGTACA  
GACAGCTGA AAGGAATGC TGTGTAAGCT GCGTGTATAT GTTTATAGAG GAAATTCAGC  
AGAAAGAGAG CAAATATGTT TTTTCTTTT GAAAGTTTT TAAAAATAT TTCAATGCTC  
TTTTTTTAA TTATATATG TGCATGTTA CAAATATGCT TGTATATCT TTTGACCTTA  
AATGCTTTT TTTATATG AGATTTGTGA CTTATTTAT TTATATATA TGTATCTGC  
CTTTTAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAA 1313

(2) INFORMATION FOR SEQ ID NO: 113:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1654 base pairs  
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

55  
50  
45  
40  
35  
30  
25  
20  
15  
10  
5

AAGAGAGAG AATACTTCT TCTTCTCTT CAAATACAG AACCTTCT CTCACATTTG  
CCTTACACT TATTTTAAA AGCTATCCTT TTCTATAGCT ATTATATCT GCGAATGCA  
TGTATGACAG AACAGCTTT CATTACAGAC TGAAGGAG CAGTCTCTA CTTAAATAG  
TTCTGTAAT TGTCTCTCTA TTATAGAAA ATTATACAG TTCTAAATAT AACTTAATTT  
TTCTTGAATT TTATTTTAC CAAATACAA ATCTTTTTT GATTTTGA GAAATTTCT  
CATAGATCA CAAATACGA CATTGATTA GATGATATTT TTCTATGAAT CCGCAAGAG  
CAGTGCAGT CCAATGCTTG GTTGAAGCTA GAAATTTTCC TGCCCTGCT GAGCTGTAA  
GCTCTCTCT CCGAAGCTTG TGAGTGGTG AGAAGATGA GAAATGTCA GATGAGAG  
AGATATACAT GAACTGCTCT GCTCTCTTG GTTCTGTCTT TGCGCAAGAG TTTTGAATA  
GCAAGGAAA TTACATGACT TCAATCTAC AGCTTTTCA GCTTTTGT TATTTTGTG  
TCTTATAT TCTCTGTCT AAAAGGCGA AGAAAGTTA CTCATATCT CTACAGAGC  
ATGCAAGCT AACCAAGTAG TTATATGCT GCTGTATGAG CAAATATAT GCTATATC  
AATCTGAGC GAAATTTCC TTCAATATC GTCCAGCTG AGTATGCT TCTCTCTCG  
CCTATTTCT TCACTCTCT GCTTACCT TGCGAGAG AGAGGCTAT TTGCTTTGT  
TCTCTCTCG TGTAGAAA AGCAGCTTT AATATATGTC TTCAAGCTG TTGGGTTGT  
TTTGTATTT TTTTCTCT CCGAAGACT CCGTGTGTT ATTGAATTT GTATTTTAT  
AATATATTT GATTTTATA AGCTTACA CAAATTTTA TTATGTGAA AGAATACAA  
GATGAGAG AAAATATTT AATATCAGC TCAATGACA AGTGTCTAG ATTATTCAT  
TGCGATCT CCTTTGTGA GCTTTTGAAG AACCTCTAG AACTTAATG TGTCAAGAC  
TTCTGAATTT TTAGCAAGTG CTATGATTT CCTGTATGAG ATTGTATAT TACTTCTTA  
TTCTTATATC CTTTCTCT TGAAATATA TGAATGTGA AATGAGGTG GATTAATACA  
AAAAAGTAT GTATATGAT AATATATCTA GTCCAGATGA AATGTGTGA TATATACCA  
TTCAAAATTA TGCATGTTG TAAATTACA GGTTAATACA AATTAATTTA ATATGCTGTT  
GATCTACTG TGTCTCTTG CTAGAAAAA TTATTAAGAG GATCTGAG TTTGGAAGC  
CAATATGATA ATATCTATG TTCTAAATGT TGCGCTATG ATAAATATTT AAGAAATG  
GATTTTAT CCGAGATAT GGTGTGAT TTATGATAT ACCAGAGTG ATGTATGAG  
TTAAATAGT TTGTAAATA TGTAAATAG TCAATATATA ACAAATCTT GACTTATTC  
CAAAAAAAAA AAAAAAAAA ATTCAGAGG GAGC 1654



5 (2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
15 GGCMAACTTT CCCCMAAGC TTGMAACTT GCAGCCGAA ACCTTGATC GTTAAAGTT 60
GGTTTGCNC GCGCCCTCG CCGAGAGAG GCGAATTGG GTTCCGCGAA GTTGGCCCT 120
CAAGCGCTCG GCAGCCAGCC ATGTCTCTCA CCGAGAGAG CGCCCTCGG CTACAGGAC 180
CTGMECTCA TCTTCTCGG CCGACTCGG CGGGGTAGG GMACTTTCA GACTGTGAG 240
GAGTGTGTG TGGACTGGCT GTTGGACTTC TTACCCGAGG GGGTGACAA AGAGAAGTC 300
25 ACACCACTCA CGCTCAAGCA AGCTTATGT CAGAAATGG TTAAAGTGT CAATGACTCT 360
GACCGATGGA GTCTTATAT CCTGTCAAC AACAGTGGCA AAATGTGGA ACTGAATTT 420
GTGGATTCC TCCCGAGGCA GTTTGATTC AGTGTAGATT CTTTCAAT CAANTTAGAC 480
TCTCTTCTGC TCTTTTATGA ATGTTCAAG AACCCAATGA CTGAGACAT TCACCCGACA 540
ATAATCGGG AGAGCTCTTA TGGGATTTT CAGGAAGCTT TTGATCACT TTGTACAGAG 600
ATCATTTGCA CCAGGAAGCC AGAGGAATC CGAGGGGGAG GCTTCTTAA GTACTGCAC 660
35 CTCTTGGTGA GGGCTTTAG GCGCGCTCT GATGAATCA AGACCTTCA AAGGTATATG 720
TGTTCAGGT TTTTCATGCA CTCTCAGAC ATTGGAGGC AGCAGAGAA ACTGGAGTCC 780
TATTTGCGA ACCACTTTGT GGAATTTGA AGACCCGAG TATGATATC TCATGACCT 840
TCATGGAGTG GTAAATGGA GCACAGTGT CTTGATGGA CATAAGAA GACAGACTTT 900
45 AAACCTTATC ACCATGCTG CTATCCGGT GTTACGTGAC CAATATGTA TTCTTATGT 960
GGCTAATGTC ACTTGTATT ACCAGCCAG CCGCTTGTGA CGAGTGCCA ACTTTAGCAA 1020
TTACTACATT GCACAGTTC AGCCAGTATT CAGGTGGCAG CACAGACCT ACTCCACTG 1080
50 GCTACGCTCG AATTAGATT CATTAAAAA TGTCTGTGCG GAAAGCAAT TCAGACAGA 1140
CAGGAGAGAA, AAAAAAAAAA A 1171
```

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 842 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```
5 GGTCTGCGCC GGAATGCAT GAGCTGCCA TGTGATCTT AGTGAATCG GTTTCGGTCG 60
10 CTCTCCGCTG TTTCCGCGGC TGGGTATTTG CTTCCGACCA TGGGCGCCAA GGGCAAAATG 120
GGCACGAGAG CGAGAGAGCA GATATTGAA GAGACAGAG AGACTCTGAA GTTCTACCTG 180
CGATCATAC TGGGGGCCAA TGCATTTAC TGCCTTGTA CGTTGGTCTT CTTTACTCA 240
15 TCTGCTCAT TTTGGGCTTG GTTGGCCCTG GCTTTATGTC TGGAGTGTG TGGGGCCAGC 300
TACCACTCTA TGAAGTGTAT GGCAGAGCA GGTTCCTCTG AGCATGGGC CTTGATGAT 360
GTTGGCTGG AGCTCAACAT GCAGAGCGC ATGGCAGAGC ACCTTAAGCA TGTGATCTTA 420
CTACAGCCA TGGTGCAGGT GCTCAGCTGC TTCTCTCTCT ATGTCGTGTC CTTCTGGCTT 480
25 CTGGCTCAG GCGCGGCGCT TTACTCTCTG TGGGTGATG TCTTGGGCGC CTGGTTCACT 540
GCAGACAGTG GCACCCGAGC ACCAGAGCAC ATGAGAAAC GGCAGGCGCG ACAGAGCGG 600
CGGAGATGA AGCGTTTATA GCGATGACA TTGTGGGCGAC AGGCCACTG CCGTGGGTGG 660
30 CTCTGTGAG GTGCAGGCC CTTCTGCTCT GAGGCAATGA GGTCTAGTC CAGGGGCGCA 720
AAGCAGTCTG AGGTATTGGG TATACTATA CTCTATAGG TCGTTGAATA ATGCGCTTAG 780
AATGTGAAA AAAAAAAAAA AAAAACTCG AGGGGGGCGC GGTACCCAT TTCNCTANA 840
AT 842
```

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1640 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

```
50 GCACAGAGGC GCGGCGAGCG GTGGGGGCGG CGCCCCCGCG CCGAGCGCT TCCCTTTCC 60
GTCCGGGAGC GCGGGTCGG GCGCCAGGGG ACCCGGGGCG ACCGAGAGCG GCGAGAGAT 120
GGATTGCGCG GCGCTCCCG CCGATCGAA GAAGAGGAA GTGATCCGAA AATCTGGCT 180
AAGTGTGCG AAGAGCGATG TCTACTACTT CAGTCCAGT GGTAAAGAT TCAGAGCAA 240
GCGCTAGTTG CGAGGTACC TGGGAATAC TGTGATCTC AGCATTTTG ACTTCAGAC 300
```

	TCGAAAGAG AGTCCATGTA ATTACAGAA GAGCAAGAG AACAGCCGCT	360
	CAATCAAAAT AACGGTAAAC CAGACTTGA ATACAAAGCT GCGATTGA GAAAGAGCT	420
5	CAATTTCGA AACACCGGTA ACCCAAGTC ACAAATCAT CTAGATTGA AATGAAATCA	480
	GACCAAGAC GAATGAATGA AACACCAAGT CAGCTTTCT GCGAGAGAG GCTACAGAGA	540
10	CTTTAGTGA TCGAGGTGA CAGACAAAT TATTAAGAAC ATGAGATAC CCAAGGCTCT	600
	TGAAAGAGTT GGTCCAGTAG CATTAGTAG AGCTTTAT CTGCTGTAC CAGTCTTTG	660
	CAGCAAGCT CTGCGCAAT CACAGAGGA GTCTCCGCTG CTGAGAAA GAGCTGCTG	720
15	TTTGACTTGA CAGATTCGA CCCCTGTGA AACCTTTAT TGTACAGAT GAGAGTCAG	780
	GAAACAGAG AACGATGCA GCAAGTAGCC AAGAAATGAG AAGAGAGCT GATGCGAGAC	840
20	ATCTGTGCG GAGCTGCTGA TACAGAGAG ATGGAATTTG AATGAGAG TCGAGATGA	900
	GCGTAAGAT ATGATCAAGT AACCTTGAC CAGCTTTCC CAGAGAGAAA TTGCTAGAA	960
	ATTGACAAA AATGTTTGA CTGCTTTTG CCGTAAAGA AAAAATGTA CCCAGAGCA	1020
25	TGAGCTTTT TAAAGAGCT AACCAATCC TTTTAAAGG TATTTTGAAT GTATATATCT	1080
	ATTATCGAA AAATCAATTT TATTTGAGT CCTAGAGCTT AAAATTTGTC TTTTGTAATA	1140
30	TGAGAGAGG CCTTAAGAGG AACGTAGCT TTTGATGCA GTTGCAATCT ACTGGAATG	1200
	TGAGCTTAC GTAAAGCAT TGTTTCCCC AACATTTTA TAAAGAGGA TCGAGAAATC	1260
	TAAATAAAT TCCAGTTTAA AATTAATGT GACTTGACTG TATTAAGACA TATTTTAAAT	1320
35	CTTTATTTAA AAGGAGAGC TGTACATGCT TCCAGATGCA CTGTAAAGAC AAATTAATGA	1380
	TATATTTCA CAGAAAAAA AAAAAAAAM MASTTGABR GSRGGRGSR ATTAABMCC	1440
40	CMAATWAS MTCSTHTKA YTTACATCA ACTGTGATCC CGAGGCCCTTA GCTTTGACAT	1500
	GCGAGGTGAG AAGAGATAG GCGAATANT TCGAGATAGA ACTATTCCTT CTGAGACGTT	1560
	GTGAGCAATT GTGCTTTGAC CAGAAATCTT AAGATTTCTT GCGTTAAAT TCACTGAGCC	1620
45	TGTGTAAAT TTTTTCCTT	1680

50 (2) INFORMATION FOR SEQ ID NO: 117:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 952 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

60 TGAATTAGN AAACACTTGG GAAACATCAT AACCTGATCA GAAAGTCCCT TTAGCGAGC 60

	TCTGAGCTT CTAGATGAGT AACAAAAAA TGAATTAAT TCTTGAAAT TAAAGATTT	120
	ATTTAATTT GCAATTTTT TCAATGTTCT AGATATCTTT AAATTTGTA TTGAGAAATC	180
5	ATTTCCTAC CAGATGACTT TATGAAAT ATTGACCTGA TGAGAGCTGA AGTAAGTACAG	240
	CTTTTGTGAG AACTTAGTG GACTTCTGAG AGATTTGAGT TGTACTTTGT ATCTTAAT	300
10	CTAAAGATAG TTTTAAATA CTCCAAAGA AAATCTGATC TCCTTTCTGA TCTAAAACT	360
	CATCTTGTGAG GTAAAGAGT AATGTGCAA AGGTGTGAC AGTTGATAG GTGAGAGGGA	420
	GCTAGCTGAG CAGCTGACT CTGCGCAATC AACAGTGA CAATTCAGCA GAAATGTAAAT	480
15	TAAATCTCC AGTAAGAAAT GCTGCGTAAAG GAAAGGGGTTA GGGTGGGTT ATTAAGATAC	540
	AGCTGCTGT ATTTAATTT GGTGTGAGG GAAAGGAGAC CTGAGAGAAA CAAAGTCAGT	600
20	ATTCCTTTT TTGAAAGAG AAAAATTAAT ATTTTGTGT CAGTAATAT GGTAGAGAAAT	660
	TGCAATGTCC CTAGAGCAAA GGAACAGTTT CAGATGAGGA GTAGAGCTG GAAACTGAAA	720
25	ATTTCAGAAA CATTGGGGGA AAGGAAATTT GCGTTTCTT TAAATGCGAG ATGTTCAGAT	780
	GAGGAGGAGG GAGCTGTGTT TTGTTGGAGT GTGTAAATTT GTATGTAGCC ATTAATGAGC	840
30	GGAATCTGCG TGAATTTAA AGGTTCGAAA AATTAATGTA AATCTGCTT TTTGTAAAT	900
	TATCTAATA AAAGCCCACT GGAATTCGAA AAAAAAAGA AAAAAGAGA NN	952

35 (2) INFORMATION FOR SEQ ID NO: 118:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

45	GAGCTCATAG GTAAAGAGC TGTGTATCCG TCGAGAGGCG CCGTGGAGCG TTGCTGAGTTA	60
	CGAGCTGTGAG CTAGAGCGAG AAGAGCTGCG TGTATCAATC AACCTGAGCTT GAGAGCGGCT	120
50	GCGAGCGCGG GAGCGAGAGC TGAATTAATGT GCTTTCAGAG GCGAGAGAGCG TGTTCAGATC	180
	ATCTTTTAC CCCAGCAATC TACAGATGTT GCGATGATGT AAGCAATGAAA GAAATGTGTG	240
	TTGAAAGAAC GATGCTGTGAG AACAGAGAGC GAGCTTAAT CTTTCTGAGC CCTTGCTCTT	300
55	ATGAGAGAGC TCTCTCTGAG TCCAGCTGTC ACTCTTTCAG AACAGAACTC CTTCAGAGAA	360
	CGTGAATGAG AAAGAGCAAT GCGCAAGAGC ATCTGAGGTT AAAGAGAGGCT AACATTTGAT	420
60	GAACTAGACA AGCTAGAGGAT GTTGAGATCCA GAGGTTAGCC AACAGAGCAAT AGAGCTGAG	480

540 GAAGAGTCCA AAGACTTTGT GGCACAAATT GCCAGTTTC AGAATAAGT TGTGTGTTTA 540  
600 ATTGAGCTTG TTGATCAACT TCGAAGAGA GCGAUAATG AAGAGATGAA GGCATCGGT 600  
660 GCTCGCACT TGTCCAAATC TATAGCAAG CAGAGAGAG CTCACAGCA GCACCTTCAA 660  
720 GCCTAATAG CAGAAAGAA AATCAGCTA GAAAGATAC GGGTTGAATA TGAAGCTTTG 720  
780 TGTAAAGTAG AAGCAGACA AATGAAATTT ATTGACCAAT TTATTTTCCA GAATGAACT 780  
840 GAAATTTTGG CTTTTHATGT AGGAGGCAA NACAAANAAA AGCCTCTCAA AACCAAAAA 840  
900 ACTCTGTAG CATTCAGCG GCTTGACCA TCACTATAT CACAGAGGT GCGGTGTAG 900  
960 GAAATCAGCC CCGTAGAGC AGCACTACAA GTCTGGGGGA GCGAGTTTTA ACATCATGTC 960  
1020 ACAGCTGTC CTGGTGGCCC TCGAGTGTAC GTTCTCACT CTTATGCTTA GTTGGAACTA 1020  
1080 AGCAGTTGT AACTTTTCAT CTTTTTTTT GTAAATTCAC AAGCTTTTGG AAGGAGAAC 1080  
1140 AATTAATTTT TGTTTTCAA TGGCTTGATG TACTTTTITT CTTGTTCCT TTGAATATG 1140  
1200 TTTAAGCTT CATGAGAGA CCGTGATTC TCTATCCCT ATGCCACAA ACAAAACAG 1200  
1256 CAGTGTGAG CAGTACCTT TAAATTCGAT CACACAGTG AUTCAGAGT TACCCAC 1256

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1143 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

30 (2) INFORMATION FOR SEQ ID NO: 119:

35 (A) LENGTH: 1143 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 GCGCCTACCA GCGCGGCTGG TCTGCTGCG AGCCGGGCGC CCGGAGTGG GCGCGCGCAT 60  
GTACTTTCCA CATTGAGTAT TCGAAGAGA GTGATCTGAA CTCGACCAAT TCTTTATGGA 120  
TACATTAAGT CAATATTAAG ACTCTGACTA CTTGACACAC TGGCTGAGC AATCATGAC 180  
GTGGAGTGG CCGACAGTGA ACGATATCCA AATACCGGTG TCATGACAG CCGGGGTATG 240  
TGGCTGACAT ATGCAATGGG AGTTGGCTTG CTTCAATATG TCTTACTAG CATTCCTTC 300  
50 TTCAAGTATC CTGTGCTGTT GACTTTTACA AATATTTATC AATATTTGCG GATGTACGTA 360  
TTTTTGCATG CAGTGAAGAG NACACTTTTC GAAACTCTTG ACCAGGTAA AGCAAGGCTC 420  
CTAACTCATT GCGAACAACAT GCACTATGGA GTACAGTTTA CATCTCAGS GAGTTTTC 480  
55 ACAATTTTC CAATATTTCT AATTTTCTCG GCAATTTCT ATAGAGATA TCAATCAACT 540  
CACTCATCC TAAACAGAG TTCTCTCTCG AGTGTACTAA TTCCAAATAT GCCACAATA 600  
60 CATGGTTC GCACTCTTGG AATTAAAG TATTGAATG TTTTGAAT TTTTGAAT GAAAAAAT 660

5 TTTACAGCTA CTGAATTTCT TATTAAGGAG GAGTGGTTAG TAACTGCAC TGTTCCTSTG 720  
780 ATATGTGAA ATGAGAGCTA TTTACATTTG AGGCGCAATG GCTGTGCTTT CAAGTGTCTG 780  
840 TTTGAAGTGC AGATTTCCAT TAAATGATGC CTCGTGTTAA TACACCTGGT ACATTTCTGA 840  
900 AGAGGCTTT TATTAAGCAG CTGGCAGGC CCACTTATA AGTTAAAGGG CATCACAGTG 900  
960 AGGCTGATGT AGATTAATTC AAGGAATTA GAGATTGTA AGAACTAGG ACCGCTTAA 960  
1020 CTTATATGA ATGGCAATG TTTTAAGAAA AGAACATTC CAGTCAATCA GCTGTGTTA 1020  
1080 TTTAAGCAG ACTTCAATGT AATCCGAAT CCTCTCTATA CAAGTTTAT AAGATTTAT 1080  
1140 TTTATTAAG TAAAAAAA AAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA 1140  
GAN 1143

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1782 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 120:

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

30 CAGGCGCGCG CCGCCGACCC AGTCTGCGT TCGTGGCGCG CCGCGCGCG CCGCCGAGG 60  
120 CAAGGACAAA CAGCTGTCA GCGACCTCC GCGGAGTGC AATTACGTG CAGCTGCGCG 120  
180 CAGCCAGAG TTCCAGATG GTTTGCGGG GCTTGGGTG TTCCAGAGAC TGGCTGTGCG 180  
240 CCGTCACTT GCTTTACAC TTGTTAGTC TCGTGTAT TCGAATTTCT GCGTGGGCGA 240  
300 TTGGCTTGG GCTGATTTCC AGTCTCGAG TGTGCGGCT GGTCAATGCA GTGGGCATCT 300  
360 TCTTGTCTT GATTGCTTGA GTGGGTCTGA TTGGGCTGT AATACATCAT CAGGTGTGTC 360  
420 TATTTTTHA TATGATTAAT CTGTACTTG TATTATTTT TCAATTTTCT GTATCTTGG 420  
480 CTTGTTAGC CCGAACCAG GAGCAACAG GTACGTTCT GGAAGTTGT TCGAACATA 480  
540 GCGCAAGTGC TCGAATGAC ATCCAGAAA ATCTAACATG CTGTGGGTT CAGAGTGTGA 540  
600 ACCCAATGA CACTGTCTG GCTAGCTGTG TTAAGAGTGA CCACTGTGTC TGGCATGTG 600  
660 CTCATCAT AGGAGATAT GCTGAGAGG TTTTGAAT TTTTGTGTC ATTTGCTGT 660  
720 TCTTGTGTT TACAGATC CTGGGTGTTT GCGTACCTA CAGATACAG AACCAAGAG 720  
780 ACCCGGCGC TATCTTAGT GCAATCTTTT GATGAGAAA CAGAGAGAT TTCTTTGCT 780  
840 ATTATGAT TGTTCATTT CTGTAATTTT CTGTAGCT CCAATTCOA GTTTAAGAA 840

GGAAACACTA TCGGAAAAAG TACCTTATCG ATATGCGAAT TAAATATTTT TACTCATCTG 900  
TTCTCTACAT GTTTTTTTCT TTCCCTGCT GAAAAATATTT TGAACCTTCT GCTCTCTGAA 960  
GCTCGCTGCG ACCTGCGAAT TTAAGTATCT CATCTCGCG CACTCTCCAG TCGTGGCTTT 1020  
CTTACGATTT TTAAGTCGAG AAAAAGCTTGG TATGTCACCA CTCTCTGCTT TAAATGTCGA 1080  
ATCTGACGCT ACATCTGACT GGTATATATTA TATGTCACAG TCGCTCTGCT AGATATGCTC 1140  
TACTGAAAAA AGATGCGGAA TTTATATAAA TCGAAAAATA TGAATCTCTG TTATGTCAG 1200  
GAAAAATCCA ATTCCCATTT TTTTCTGGTC TTTTTCAGGA AGATCTGCTG TGTATAAAAAG 1260  
TGTATAGTAA AAAATATATA TTMAGTCGTA GCTCTTTATG ATMACACCA TGTATCTGAG 1320  
AAATAGTTAT GCTTACGAAA ATTCTGCTTT AATTTTTCAG TTTTACAGCT AAGTCGAAAG 1380  
GGAAGCTGCT TTAATGAAAT GTTCTAATCT ATATATACAT TTAAGTCGAG CCGTCATCGAG 1440  
AATGACACGA GTTTTACGTA ATCGAGAACT ATATCTAATAT GATCTTGATA TTGTTTATTA 1500  
ATAATTTGAA GTCTAAAAA CTGCAATTTT AAAACAGTAA GTATTAATCG GTTGGCCGAC 1560  
GTACAAAAA GATATTTGAT TATCTTAAAA ATGTATATAT ACCTCTTCA TGAATATCTT 1620  
CAATATTTGA ACACGAACTT GTTAAACCTA AGCAATTTTG AATATGATCT CCGATATTTT 1680  
GAATATGAAA TCGATTTCTG TGGCTCTGTA TATCTGCTTA AAAATTTTAA GAGACGAAAC 1740  
CTTCTCTTCT GTATGCAATG TTGAATATTA AGAAATATAT GG 1782

35 (2) INFORMATION FOR SEQ ID NO: 121:

(1) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 610 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GTTCGCTGCA GATTCTGCTG GCTTTCTGAG CCGTCTGCTC TTGCGCAAGA TCGTTCAAAA 60  
TATATATAAA AACATATGGA TCCCATGAAA GCGCTACTAC ACCAAAGTTT ACCAGAGAAT 120  
TTTGATATGAGA ATGGGAGCTGA TGGGCTTCAAT CATTATATAAA ATCCGAGCTG CTGATATAAAG 180  
AAGTAAAGCT TTGAAAGCTT GAGCGGCTCG TCGCTGTCAT CACACACAGA TTTTACTGCA 240  
GTATATGTA AAAAATACCT GATCTGCTCT GTTAAATTTCA GCAAGCGCTG TTGAAATGAGG 300  
AGCGTGAAC GTCACTGTAC ACTATATATA GTACCGTTTA CTTCATGCGA TGAATTAATG 360  
GATCTGTGAG ATGCACTGCT ACCTGTGACT GCTTTCAGAG TGTTCCTCT CAGCGCTCGG 420  
GCGTGTGAGG CATTACTGTA GTATATATTT TTGTCATGAG CCAATGCAAT CAGATATCA 480

CTGAGCAAGC GATCATTTGG AAAATATATG CTCAAGTCTA CAGAGCTGAG TGAATGAGAT 540  
CGAAGCACTC ACTGTATATC TACTGTCTAA TAAAGTTTAA TTAACTTGA AAAAAAATA 600  
AAAAAAAA 610

10 (2) INFORMATION FOR SEQ ID NO: 122:

(1) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 526 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GATACGCTCG CAGATACCGG TCCGAAATTC CCGATTCGCC AACGTCAGG CCAAGCGCTC 60  
ACCGACGCTT CCGGCGACCG GTTCGAGCGG AACCGAGATG GTCAAGATGA TCAAGCAAGCT 120  
GAGCTGCGCC ATCTTCCACA ACATCTGCGG CGTCTGCTCT TTCTTCTCTG TCGTCTCTTA 180  
TCACTAGTGG GCGCTGACCA ATCCGACAAA CCGAGAAATGA AAGTGCAGCT TTCTTCGCGCC 240  
CAGGCTTCCA GAGCATATAT TGAAGCAGTA TGAAGGATAT GAGGCGGCTT CACACTTCAC 300  
TTGATCCCTT CTACCCATCA CAAATATACA AGCAATACCA CCGGATTTT TCCAAACACAC 360  
TTTATATTCG TGAAGATCTT CTTTATATCT ATGAAACAGG AAGCTCCGAC TGAATATGCGC 420  
CGAGTATAGG GCTTCTGCTT TCTACTCCCT CCCCCAATA TAAATATATA GACTTTTAAA 480  
AAAAAAAA AAAAATTTCC NCGCGGCGCC GGTACCAATC CCGCTTA 526

40 (2) INFORMATION FOR SEQ ID NO: 123:

(1) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 2081 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

TGTACGCTTC CCGAAATTCG CCGGTCGACG CAGCTGCTGCG GGGGAAATAG CCGGCTTCGCG 60  
AAGCCGCGGCT CCTTTCGCTC CCGAATAGCG GCGCGCGGCGG CCGCGCGGCGG CCGTCCGAGCA 120  
GATGCGCGCT GCTCTCTGCT TTCTCATGCT TCGCGCGACG CTGTCTGCTG CCAACAGCGGG 180  
CGGCTTACGA GGTGCTCATC CAGAAATTCG TCAAGCTGTA CCGGACACAG ATCGACATAC 240  
ACCGAAATTT GGTGTGCGAG CTGTGCGCGG AAGATGTGCG CAGATAGTGG GACTTCCCA 300

5 AGGGCTTCC GSTRAGGAG CGCTGGAAG TCGCTCTGT GCGTTCCAG ATCCAGCTCA 360  
CTACCTGGG AATCTTACA CTTTCAGCA CTGTGTTTTT CTGCTGTAT ATCCAGGAA 420  
GCTTCAGACC AGCCATCAG TATTTTGGG ATATTATTAG CTTGGGACG AGATTGTTC 480  
AAGGGCCCG GATTTTAGA ATTCTGTTA TTGTACAGA AOMTACCT AAGGTCTTG 540  
GGAGAGCGT TCAGAAATTT GATTTAGAG GTGTAAACT GTTACTTCCA AGACCAAGT 600  
TTTCAATGCT ATTACAGAA GTAGAGCGG CATTACAGA GATTCCCGA GTACAGAGT 660  
TTGTATTATT TGGAGTAGA ACTATGTGT GCATCCACA AACTGCCCTG GAGCTAGTTG 720  
GCCAGAGCT CGAGTTTCAG ATTGTGCTG ATGCCACTC ATCAGAGAGC ATGATGACA 780  
GGATGTTTCC CTTCCAGCGT CTGCTCTAR CCGCATCAT AGTCACACG AGTCAGGCTG 840  
20 TTCTGCTTCA GCTGTAGCT GATPAGACC ATCCAAATTT CAGAGAAAT CAGATCTTA 900  
TTAAGGCGAG TCGTCCAGAG TCGGTCTCC TTTCGAAAGT ATAGGACATT TCGAGAACTG 960  
GTATGCTACT CACTGTGAA GGACAGTCAG GTGAGGACT GTAAGCCAC ACAAGCTCTT 1020  
25 CTTATCTCTA CTGAATTAA ATGTATAGT CAAAAGGCG TCTTTTTTG GCGCTCTTAG 1080  
TCAAACTTAA CAGCTAGAC CATTGAGTA CCAGCTTTA GTTACAAAG TCGAAGGCTT 1140  
30 CCGGTGTCG TTACCTTCT TTTTGTGTA TGTGCTTTA TTTATTAAA AAAATTACA 1200  
TCAGATGCC TGTTTTCTCT CTACTGTGTA CTCTGATCT ATCTTTCCA AGTCGAGCT 1260  
CTTGAGAGT TTCTTAATTT TGTTCACCTT AAGAAATG AGTCACAC ATGATTTGG 1320  
CTTTTATTT ACTGTAGAT GTTATATGT TATGTGGAT GTATGCTTT TACTTTACG 1380  
ATTGATGGA ATAGATATT TGCATATGA TTACCCACA GACTCTGAA TCAATTTAC 1440  
40 CACTGCCCTC ACATGTTTGT CCACTTAGTG ATTTGCTTG ATCTATCGT ACCAATGAT 1500  
GTTGATPAT TACATPATTT TCTTGACTAT ACTGATTTCT TATTTTGGTC ACTATTACTA 1560  
45 AMTCTGTTT AATATTCTCT GTTTTAACTG AAAGGGATG GATPAGAGG GTTTGCAATG 1620  
CCATATTATT GGTGAGGCG TGTTTTACA TCTTTGAGT ATGGCTTGT GATATCTTT 1680  
ACCAACATCT TGAATPATTA TTCTAGTCT CACAGATTT ACAGAAAGA TAAAGCTTGG 1740  
50 GTGGATPATC ATTTTAAAT GTTCATGTTT TATTCATPAT TTCTTCCAC TACTCTCAA 1800  
ATATTGTAT GCAAAAGTC TCAGTATGA TTGTGTAGTA TTAATTTGT GGTCAATGTT 1860  
TCTCTGAT AATTTATTT TCAATPATTA CTHTTAGAG GGTTTTGAA TGTTTTCAA 1920  
55 ATATGCGAA TGTGAACTG CTGTCTTTTA TATTAAAGTA ATTAAAGAA ATGTATTGTG 1980  
ATTGAATTA TTTTGGCTC CACAGATG CTCTATGAT ATTTCTCAG GATTTCTPAT 2040  
60 ATTATTATA GGTATPAAA TCTTGACATT TATATCTTT C 2081

## 5 (2) INFORMATION FOR SEQ ID NO: 124:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1717 base pairs.

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

15 CCGCGCGGA CCTGACCCG CGGTGGGCTA GGCGAGGCG CGAGCCCGG GCGCGGAGC 60  
TGTGGATCT TCAATGATGAG AGATTTTGGG ACACTTCTCT CTCTGTGTG TAGTTGATAG 120  
TTTGTGTGT AGAGATGCG TGAAGTGTTC AAAACCTTTC TCGAGAGCT TCGCAGAGGA 180  
ATCAAGACT CCATCTGGG TATTGTGAC ATCTCAAGC TAGATGCTCG AATCCAGCA 240  
AAGAGAGGG AGCAGGCTCG AAGAGGGCA AGTAGTGTCT TGGCAGAGAG AAGAGCCGAG 300  
25 AGTATGAGC GGAAGCAAGA GAGTAGGCA GGTATGTGTA GTAGAAATTT CAGTGTGTT 360  
CGTTGGAATG GTGAGTGTCT CTGCTGAGT CTCTCTGTGT TTTATGAGT ATTATGCT 420  
GTCTTCAAGT CGGTACAGC CCGAATATC GGTGACCAT CACTACATGG AGATGTTTG 480  
TCTGTGCTG AATTTCTCT CAGTCAAT TTCAAGTCTC TTTGAGTCT CCGCTGTTT 540  
GTCTTAGCA AAGTGTGAA TCCATTTGG TTTCAAGATA TAGCTGACT GGCATTTGAG 600  
35 GTATCAGGA GAGAGCTCA CCAATTTCT AGTGTGACA AAATATTC TGACATGCTC 660  
TTCAAGCTT TCTGAGGC TCTTTTCTC ATTCAGGAA TGTTTGAG TCTCTTTCCC 720  
ATCATCTTC TCGTCACT GTTATGCTC CTGCATATGT CCTTCTCTA CTCACGTGAC 780  
40 TCGTTGAT ATCTTTGCTT CATTAAAGA ATTGAATGC ACCAGCGGT GTCTAACATA 840  
GAAGGAATT GCGCTTACTA CTTTGGGTTT GGTTTGCGCT TGGCTTTTCT CACAGCAATG 900  
45 CAGTCTCAT ATATATCAG TCGCTGCTT TTCTCTATCC TCTTTTCTTT ATTCATATC 960  
AGCGCAATG AAGCAAGAC CCTTGCAAA GCTATCTCT TCGAGTTGGG CCTCTCTCC 1020  
50 TTGTGTGCT TCTTAAGCA CAGACTCTTC CACAGAGAG TCTACTGCA GTCGGCCCTG 1080  
AGCAAGCTTA CTCTGCGA GAAATTTCT TCAAGGATC GTTGGCTTC CAATCTGAG 1140  
GCTACTGCG GTCACTGAGT TCGCTGCAAT CCAAGGGA TGGCGCGAT TCGAAGAGC 1200  
55 TGTGGAGCT CTTTTCCTG TTCACTGCC GCTGCGAG GAGGAGGA CCGCTCTCC 1260  
CAGGCGGCT CTGCTATTC CTTTCTCT GAGGATGCA AATTTTCTC TCTGTGTC 1320  
GTAGGCGGA ATGTTTCTG ACACAGAT GTGATTTT ACATCAAGC TGAGTCTGAA 1380

AGGACACAG GTTTCTCTG ACCATTTTC TACAGTTGC GAGTCCCTCT GCTCGAGTC 1440  
ATTGAGACAC TTGTGTTTTC TCCCTGTGC CAGCTTTGC ATCTGCTTT 1500  
CTACACCCCT TCGATGATG GATTGTGTA TTCTAGCTCT TGTATTTTC GAAATTTGTA 1560  
ATTGTGTTT TTCTCTGTA AACATATCA TTGATATGAG GAGTAAAGG AGTGTCCAG 1620  
TTCTCTCTG TCACTCCCTT TATACCAAT ACTGTCTCT TTCTGTAC TCAAGTTAGG 1680  
TTTGTCTCT TCTTGTCCA CTCGAAAAA AAAAAA 1717

15 (2) INFORMATION FOR SEQ ID NO: 125:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 804 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

CGAGGCTTCC GGTGCTATG TATGTAGAG GAGAGACCC TCCCGCAAT TCTGTAGCT 60  
TCTTAGTTC GACTAGGCA GTAGCCCGAG GACTCTAGT GCGCGCTTC AGTCACTGC 120  
CGCTGACAG GAGTGCCTT CCGCTGTTT GCTCTCTTG TCGCGGAGG GCTGTGCGA 180  
ACAGCTCAC AGCAAGTGC AGAGATATA TTGTTTTTC ACTTACTGA TTATGAAAT 240  
ATCAACAGT TTGTGTTTT TATCTGCGA ACATATCCAT TTCTGTAGAG TATGTAGCA 300  
TCTGTACTT TTCTTATCC TGAATCAAT GAAATGCGAG TATGTAGCT CCTAGGATTT 360  
GTCAAGATG GAGAGCGAG TGCATCTTC AAAATTTGAG GTCTTAAATC TCGAGAGCA 420  
AGCCAGATC CTTTGTGAGC CATGATATTT GTCCGAGTC GATCTGTGC TCGATTTGCA 480  
ATTGAGTGC AATTATTTAG CATATAGCT CAGCAGACT CTGTAGTGA TCTGTCTGA 540  
TCTGTAGTTC ACTCATCAC TCAATTTCA CAAAGATGT TCGACATTT CTACATTTT 600  
GCTTCAAT TCTGTCTTC TCGAGCGAG ATGACAGCA GCGCATTTGA AATGTCTAT 660  
CGCGAATG TGTCTTTGA ATGTATGAA AACTTTCAA GAGCATGAG AACAGACCT 720  
TATTTTTCAT AAACATATTT TGAATTAAT AATTTTAT GAAATTTGTA AAAAAAAA 780  
AAAAAAAAA AAAAAAAA AAAA 804

55 (2) INFORMATION FOR SEQ ID NO: 126:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 431 base pairs

60

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

GGCAGACCC AGGCTTTTA AGCAGCTCG CCGTAGAGAG GGGCTGTCT GCGAGCTTG 60  
GAGGGTCTG GAGTGGGGCT GCGCTGTAG GCGCTGTAT GAGTACTTT GCGAGACT 120  
GCTGGGGTGA ACTTATTTT AGCCTTCCC TTGTCTCTT TATGTAGAA CAGAGAGAG 180  
GTGGGCAAT CATGTATCT ACATGTAGG TATTTCCAG CAGAGGGCT TCTGGAGAA 240  
GGCATGAG GATTTCTTT CAGGAAATG GTCCATTAAT TCAAGCGAA GCGATTCTAT 300  
TAAGTTAAT CCGAGACTT TGTGCGCGAG CTCTGTGTA TTAGGGCCC TTGGCGAACA 360  
CTTCAAGAG GGGCGAATAA GACTTTAG TTTTTATGTT TTAACAGAGG AACCGCAAA 420  
GGATTATTT G 431

25 (2) INFORMATION FOR SEQ ID NO: 127:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3752 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

AGGACAGAG AAGATCACT GACTCAGAA CTAGATAT CATATGACC AAGACAAAT 60  
AACTTCAG TTCTTAAACA TAAAGATTT CAGAGACTC TTGTGTGCAA GCAAGCTGT 120  
TATGATACCA CATTTAGAC TCGCAGAGCA CTGAAGAAA AAGCTTTGCT TCCGAAAGAT 180  
ASTCAGAAC TTGACAAAT CTAGAGAAA GTTCAGACA TATGTAGTAC TGTTTTGTGC 240  
AAGTCTGTG ACGGACAGCA CAGATTGAG GAAAGCTGC TCTTTTGGG TCAATTCATG 300  
GATGTCTTC AAGCATTTCT TGACTGTGA TACAGATGAG AACCAAGCT GCGTAGAGAC 360  
CAGCCCTTGC ACCGGGAGAC TTGACTCTT CATTAAGTC ATGTATGAC ACAAAGTTTT 420  
CGAGAGGAA CTGTAGAAA GAAAGAGAA CCGTTCAGCT CTTGAGAGCG TCAAGCGAG 480  
ACGTATTTGA GATATGTGAA GATGACAGCA CTGTGTAAA AAGACAGCTC CAGGACTGA 540  
GCACTGCTG GAGCAGCTCT TGTAACTCT CTGTTTTCAA ACAAAGCTCG CTGTAGAGAG 600  
CTTTAAACA ACGGAGATG TTGTAGACA CAGTCCAGT GCTGTGTAGG TGGCTTTCTG 660  
AACAGAGCA AAGCTTGC TTGTGGGAG GACTCTGAG ATGACAGAA GCGCTGTGAG 720  
TCTGTATG AACAAGTGA GAAATCATG AAGAAATGAG AAGAAAGCG AGTGAAGTT 780

840 AACTCAGGAG TAGCCATGGG AGAAGTATC CTGGCTGTCT GCGACCCGCA TTGCATGCA  
900 ACCATCAAC ACTGATATC CATCTGCA GCTGCTCTG AGGAGGTCTT GACATGGGCT  
960 AAGCAGACC AGCAGCTCTT TGGACGGCC TTGTGAGAAC TGGTGGCTAA TCGTGAGCTC  
1020 CTGGAGAAC TTCTGGCATG GATCCAGTGG GCTGAGACCA GCTTCATATCA GCGGATCAG  
1080 GAGCCATCC CCGAGACAT TGAACCGATT AAGCCCTTAA TCGCTGAGCA TCGAGCATTT  
1140 ATGGAGGAGA TGACTGCGAA AGACCTTGAC GTGGACCGGG TCACGAGAAC ATACAAAGGG  
1200 AAAAAATAG AGCCTACTCA GCGCCCTTTC ATAGAGAAAT CCGGACCGG AGGCGAGAAA  
1260 TCCCTTAATG AGCCAAACCC TCTTCCCATG CCAATCCCTT CACAGTCTCA AGCAGAAAGC  
1320 CCAAGATCA ACCAGCTTTC TCCCGGCTGG CAGCAGGTGT GCTGTATAC ACTGAGAGCG  
1380 CAAGAGAAC TGAATGATGC CTTCGATCGG CTGGAGAGAT TGAAGAAAT TCCCACTTT  
1440 GACTTTGAT TCTGGAGAA AAGATATAG CTTTGATGCA ATCAGAAAA GTCTCGAGTG  
1500 ATGCAATTTT TCCGGCCAT TGAATAGGAC CAGGATGGGA AGATACAGCG TCAGGAGTTT  
1560 ATCGATGGCA TTTTAGATC CAGTTTCCC ACCACAGAT TAGAGATGAC TCGTGTGGCT  
1620 GACATTTTGG ACCGAGATGG GATGTGTAC ATTGATTAAT ATGAAATTTT GCTGCTCTTT  
1680 CATCCCAACA AGCATGGTGA TCGACCAACA ACCGATCGAG ATAAATGCA AGATGAGTT  
1740 ACAGAGCAG TGGCTCATG CAATATGCA AAGAGTTTC AGGTGAGCA GATCGAGAG  
1800 AATAAATACC GATTCTTCTT CCGGATATAG TTGCGGATTT CTCAGCATTT GCGGCTGCTC  
1860 GGTATTCTGC GCAACCTGTA TGGTTCGGST TGGTGGAGA TGGATGGCTT TGGATGAAT  
1920 TTTAGTGAAA AATGATCCCT CCGGAGCAG AGGTAGACT AACATTGAC TTAGAGAGAA  
1980 ATTCAATCTA CCGAGGGAG CATCCGAGG ATGACCCCC TTCCGCTCAC GGGGTGAGAG  
2040 GTCCAAACCA TCTTCCCGGG CAGCTTCCC TACTGTTTCC AGCTCCAGTG CTAGTCAAG  
2100 TAACCAAGC TTGATATCCA TCGCATCTTC TCGACCGACC CCGACCATG GAAACAGGT  
2160 TATCCCATCA TCAGTAGACA AGTTGAAAG ACCAACAACA ACTTTTCAAT CTAGTGGAC  
2220 ATCCCTTCTT GGTGATACA GCAATGATT CTTCGCGGC CTCACAGGT GCCAAACTA  
2280 ATCGGCGAGA CCTTAAMAG TCTCCCATG CCGCTGGGAG TCGGCTGGG AGTCCAGCCG  
2340 GAGTGTGAG CAGCAGCCGG CAGGAGAGT AGCTTTCTGA CTTTGACTTC TTAGAGACCC  
2400 ATTGCTTTGT CCGACACTTC AGAAGCAGC GCTCGAGGGG GCGAGGCA CTCGAGAGA  
2460 GGGCTAAACA AAGCTTCCA AATCCCAACC ATGTCTTAAGA AGAACACAC TGCCTCCGCC  
2520 AGGACTCCAG GTCCGAGG AGTAACACT CTAGACACC CCAAGCACT ATCCACTTGG  
2580 AATCTCTGTC CATACATGG GTGTATTAAT ATTCTGAGG GAGAGATTA TATTGTTAAA

2640 AGGTAAAG AATATTTG TTATGAGCT GCTTATTTT TTTTCTTTT GTAGTTACT  
2700 ATTTCTATGT GAATATTTAT GTAGTAAAA TTGCTCTCT GGTAAACCTG TAATGATGG  
2760 GGGCCAGAAA TGAATATTTT GAGAAACCA ACTGAAAGG TCAGATATCA AATGTGTAT  
2820 AAAAAAAA AGCCTATTA ATAGGTTTC TCGCGGTGC AGGTTGTAA ACCTGCTTTA  
2880 TCTTTTGA TAATCTCTAA ATGATCTTC TTATTAAC TGACTGCTA TCTGAGCAG  
2940 ATAAATATA TTAATAAAT AAGATCTCG CAGTGTTTA GGAATCTTT TTTGTAAAT  
3000 CAGCGACCC TCATATGCA AGRACAGG GAGGGCTTT TTCCATTTT TAATGTTTG  
3060 TGAATTTTAG CTAAAGAG GGAACCTCAT CTAAATACA TTTCACATG ATACAGCAA  
3120 AGAGTTTCA TGCATACTG TCTTTGATA TTGTTCACT ACTGGGTGT TAAAGACAA  
3180 ATAGCTGCTA GAATTCAGG GTAAATGTA GTGTCAGAA AACGTCAGAA CATTTGGGT  
3240 TTTAAACTGA TTTGTTGCTC CCTATCCAG CTAGACACA GTAACTCTTG TGTTCACAG  
3300 GAGCCAGACC CTGCGCAGG GATAGGCTCG TTGTGACAT TGTGAATTC AATTTGTTT  
3360 TATCCACTTT TTTTGTAT TATTAAATG GTGATCAG TCCTCAGAAA CTGAGGATG  
3420 AATTCACGA GCGTGTCTG AAAATGTGA GTTAGACAA ACAGTGTCT GTCTTTTAAT  
3480 GGAATTACCC AGCACTATG TTACCATGC CTGTTGCTT TGTCTTTT TTGTGCTTA  
3540 TAAGTTCAAC TCACAGAGT ACCATGAAA GGGGCTGCT GGGGCTCTG TTTTITAGCT  
3600 GCTGTTCTTC AGCTCCGACC ATGTGCTGT GTGATATCT CAATTTGTTT TAATGAGGC  
3660 AGAACTGA GCTTCACAA TGAATGTTT AGAAGCAGA CACACTTTG TATTAATAAT  
3720 GCTTGCAGTA ACAGAAAAA AAAAAAAA AAAAAAAA AACTCGAGG GGGGCGCGGT  
3752 ACCCAATTC CCGTATATGA TGTAAACAA TC

(2) INFORMATION FOR SEQ ID NO: 128:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1144 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55 TGAACCTCTG CCTCCGCGC TCAGTGTGAG AGCTTTCTG TTTTGTGCA GTGCTGCTC 60  
GTTAAACCA GGGGCTCTG GTCCACCACT CCAATTCAGA GCTCCATTTG GTCCAGAAC 120  
CTTAGCAGG CTTTCCCTTC ACAGTCCAG CAAACAGCT GCGAGCATG GGCCTCATG 180

GAAGAGAAAA GTCCCTGCT CCTGACGTC CCTGACGTC GAGTCCAGCT GAGCACTTGG 240  
GAGAGACCTA GAAACCTGG GGGATTGGA GCGCCAGCAT TGGAGAGGA AGCTGGACCC 300  
CTTGGCTTGA TTGCACTTT CCTGACGGA ATTAAATTGA GCAAAAAACA AGCAACCTTA 360  
GTGGGCGGTC GTCTAGATCA TGAATGCGA GTTCTGAGA CAGCTTTTGA AGGCTTTAC 420  
TGGACCTGCC GTCCCGACCC TCCCTCTCT TCGAAAAAG ACCGAGCGAG GCGAGCTTCA 480  
GAGCACTGCC TTCTTTGAGA TCTTCTTGC AATTGAGTA AGATGAGATT TGGTGTGTG 540  
CTTTAGAGGA AAGTCTGGA CTGACGACA AAGCTTAATA AGACCTGAC ATCTGAGAC 600  
CTTCCCGTT ACTGCTTTT GAGCACTGT CTTCGCGATG CTTTATTAT CTGTATGAC 660  
AGCAATTGA CATTACAGCT AAGGAAATTA TTGAGTGA TTGAGAAATC CTGGGATGTG 720  
AGATTTGT TAAATTACA AGTTGGTTT TTAAATTTT CTCAATTGA TGGCCAGGA 780  
TCTAATTTTA AAGCTATAG AGCACTTTT GCTAAAAA GAGTATTTT TTAAATGAG 840  
GCTGCTTGA TTAAAAACA GATTACAGA ATGTAGATCA ACTTAAAGAC TGTAAATGAA 900  
TGTAAAAA TTGATGAG AGCAATGCA TTTTCTGTC TGTATTACT TGAAGTATGT 960  
AAGATTGTA TACCTGACT TATTTTAAG ATGAACTGAA ATGCAATAG CCAAGCTTG 1020  
AGATACAGA TTGAATGTGT AATTCTTAA AATACACTT TGTGTGTAC TTTGAAATTA 1080  
ATGATGCTTT TTGAAAAA AAAAAAAAAA TCGAGCGGG GCGCGTACC 1140  
CAAT 1144

(2) INFORMATION FOR SEQ ID NO: 129:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1830 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

GAAGAGAG GAGACCTTG AGCTGTTC TGAACAGAG GCGCATSTG GAGAGAGCC 60  
AGCGTGTCT GCGCAATGCG ATGATGCGAG CAGCGACAT CATGCGAGAC AGGCGCGAG 120  
GATGAGATT CTGCGCAAAA AGCTGCGAT CAGAGCGAG ATGCGCGAG GTGCTCGCG 180  
CCTTACAGC CTCTGCGAG CCGCGGTGGA TGGGAGCTC TGAACAGCA AATGCCCGAC 240  
GCGGCGCGCG GCGCTGGA GCTGAGATT GAGAGAGAC AGCAAGCGAG GCTGCGCTTC 300  
TCCAGGATG GCGGAGCT TCCGAGAGC GCGTTCGAG GAGCTGCCA GCGTCTGCC 360  
TGCCTCTTC GCGAGAGAG CTGCGACCC TCGTGTGAT GAGAGCTGAG CTGATTGCGC 420

CTGGGCGGCG CAGTCTTCA GAGCTGCGC AAGAGCTGA GGGGTGCTG AGACCATAG 480  
AGCTGATGAG AGCACTGTGT GCTTGACTT GAGCTCTGC GTCCCGAGA CCGAAGGAA 540  
GTGATGAGAG GCGCATGAG GCGACCGCG TCCCTGGGA TGGCTCCGCT GCACTTTGA 600  
AAGCGGTTT TCTTCAGAG TCCAGTTCC AGGTGACAG AGGTGCTGC TCCCTCTCAT 660  
CTTACTTCC AGGTGACCC TTAACCTTGA CTAACTGCT TGTGCACTT GAAAAAGCT 720  
TGCCTGTG GCGAAGAGAG AGAGCGGCG TCGATGAGC CTGTACAG AGATATCCCG 780  
AGAGCGAGAG GAGCTGTGA GATGACCGC CGTGAAATG GCGGACAGAG CCGTGGGAG 840  
GAGATTGCT CTGACTTAT TTATATGCA TGAATCTCT GATTATTTT GGAATTTTT 900  
GTGTGATG TTTGCAAGT TGTTTTTT TAAAGTTG TATATATTA TTTGACATTT 960  
TACATTTCA AAGAAATAT GTGTCTAAC AGCGACGAA CAGAAATG TATTGAGAC 1020  
TGTCCGCT TCTACTAAA AAAAAAAAAA AGCAATTTA TTGAAAAAT 1080  
AAAAAAAA ATGTGTCTT GTTTTATAT ATTAGCTTG TAAAGTGTG TTTTGAAGTA 1140  
TGAAGTGA TTTCTTCCC GAGCTCGAT TCTGAGCGG TTGGCGGTT TTAAATCTG 1200  
CTCTTTGA GTGCTGCTT AGAAGGCTT GAGAGTGTG CAGCGAGGT TGGAGCTGA 1260  
GGGATGAG GCGCACTGA GGTGCGTGT GAGAGGTTT AGAGGTGAC TGGGAAAGC 1320  
CGGGGAGT GTCGTGAT TTAATTTCT GATGTATCT TCTGAGACA GGGCAATTG 1380  
GAGATGACA GCGAGAGAC AGAAGCGAC CCGCTGCTT GGGAGAGAG GAGCCAGCA 1440  
AGCGCTCG GAGATTTCT CCGGACCTT GCGATGCGT TCGAGGAAAG GTGAGAGCTG 1500  
GGAGGAGAG GGGGAGAGC CTGTACCCC TGGAGGTG TGAATTCAG TGGGCGCTCC 1560  
CTGCTKCCC GAGGCTGGG AGCTTGAAG CCTCCGCGA TCTGCGATTC GAGCTTCCC 1620  
CCTGCGAGG TCGCTTCCC TCTGTGCGG CAGGATGAC GAGGCGAGAG AGTGGCTTG 1680  
TCACTGTATC TTGCATCAG GACAAAGAG GAGCGCTTT AGCGTCTG CCGGAAATG 1740  
GAGATGCC AGGCGAGAG CATTGTGAG TGTATTACT TAAATGTAC AGATTTCT 1800  
CGTAAATTC TTGATGATT TTTTATTAT 1830

(2) INFORMATION FOR SEQ ID NO: 130:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1864 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60



(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

5 GGCCGCCCGG ATGGCGACCC CAGCTCTGG CCTAGACACA GGGCTCTGG TGGAGACTT 60  
TGTAGTTTAT AAGCTGAGG AGAGGGTTTA TGTCTGTGGA GCTGGCCCCG GGGAGGGCC 120  
AGCAGCTGAC CCGCTGCACC AAGCATGCG GCGACGCGA GATGAGTTG AGACCCGCTT 180  
CGCGCGACCC TTCTCTGATC TGGCGCTCA GCTGCAATG ACCCAAGCT CAGCCCAACA 240  
ACGCTTCACC CAGGCTCTCG ATGACTCTTT TCAAGGGGGC CCGACTGCG GCGCCCTTGT 300  
AGCCTTCCTT GTCTTTTGGG CTGCACTG TGCTGAGGT GTCAACAG AGATGGAAC 360  
ACTGCTGCGA CAGTCTCAGG ATGCAATGTT GCGCTACCTG GAGAGCGGC TGGCTGACTG 420  
GATCAGAGC AGTGGGGGCT GGTATGCCA GATCAGTGA GCTGAGATGG CTGATGAGT 480  
AATTTGCACT GAATTTTAA GCGACTGTGA CTCTGCTGCA AGTCCCGCAG ATCTTGAGGA 540  
GCTGCAAGCT ATCAAGGCTC CAGTCAAGGA GATGAGGAA GAAGTGAAGA AGCTTAAGGA 600  
GCTACAGAC GAGGTAGAGA AGCAGATGA TATGAGTTCA CTTCAAGGCA ATGCTGGGCC 660  
GATGATCATG TCAATTTGAG AGAAGATGGA GCGTATGCC CATTCACTCT ATGTTGCGAA 720  
TGTGCACTAT GGTGCAACAG CAGAGAGCT GGAAGCTCAC TTTCATGCT GTGTTCACT 780  
CAACCTGTTT ACCATCTCT GTGCAATTT TATGTGGCAT CCGAAGGCT TTGCTATAT 840  
AGATTTCTCA CACAAGAGT CAGTGAAGC TTGCTTGGCC TTAGATGAGT CCGTATTAG 900  
AGCAAGCGAA ATCAAGTGA TCCCAAAAG AACCAAGGA CCAAGGATCA GCACACAGA 960  
CGCGGGTTTT CGAGAGGCC GGTACCGGC CCGGACGACC AACTACAGCA GCTGCCCTC 1020  
TCAATTTTAC ATGTGTTTAA ACACAGGCC CCGGGTCCG GTCTACAGG GCGCGCTAG 1080  
AGCGACATCA TGTATTTCC CTTACTAAA AAGTGTGTA TTAGAGGAG AGAGAGGAAA 1140  
AAGAGGGA AGAGGAAA AAAAAAGAT TAAAAAATA AAAAAAATA AAAAAAATA 1200  
ACCTTATGAT AAAAAAATA TTTTAAAAA AAAGATATA CTGTGAGG GCGGAGATC 1260  
CCTAACTAA CTGTGAGGA GCGACTGCT TTGGGGAGTA GCGGAGGCC CAGGGGATG 1320  
GGCAAGGGGC TGTCTATCA CTCTGGGAT TCGCATGGA CACGTCTCAA CTGCGCAACT 1380  
GCTTCCCAT GTTTTCCCTG CCGACCCAC CCTCTTCTC GGGTCCCTG CCGCTCCAGA 1440  
TTGCTGCTG ATCATATTTG TTTCCTTTTG TGTATCTTTT TCTGTATTGA GTGCTCTTCT 1500  
TTCCAGTTT CTGTAGCCG AGATCTCCG TTCCGCTCCC AGCGGCTCA GTGTAAATTC 1560  
CCCTTCCCCC TGGGGAAATG CACTACCTG TTTTGGGGG TTTAGGGGTG TTTTGTGTTT 1620  
TCAATGTTT TGTATTTTG TTTTATTTT TTTCCTTTG CTTTCTTCC TTTTATTTG 1680  
AGCGAATCG AGGAAGTGG AACAGGGAG TGGAGTGG ATTTGTGTTA TTTTATTAG 1740

TCATTCCAG GGTGGGAAT TTTTTTTTA TATGTGTCAT GAATAAGTT GTTTTGTAA 1800  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1860  
5 AAAA 1864

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2041 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 131:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

20 GGCAAGAGG CCGCGAAGG CCTGAGACC GCGCGCTCC CCGGGATGCT GAGCAAGCG 60  
CTGCTGGCC TGTGTCTGC GTCAACCGC AGGAGGATGA AGCTGCTGCT GGCATGCGC 120  
TTGCTGGCT ACCTGCGCTC TGTTTGGGC AACTTCTTTA ATATGAGTC TATCCAGAA 180  
AATGCTGAC TAAATTTGA AAGCAGATTT GAGAGATGG TTGAACCTCT AGAGAGAA 240  
ATCAGAGATT TAGAAAAAG CTTTACCCAG AAATACCCAC CAGTAAAGTT TTTATCAGA 300  
AAGCATGGA AAGAAATTTT GATACAGGA GCGCGAGGT TGTGCGCTC CCACTACTCT 360  
GACAACTCA TGAATGAGG CCGAGAGGT ACCGTGTGG ACAAATTTCT CAGCGGCGG 420  
AAGAGAAAG TGGAGCACTG GATGAGCAT GAGAGCTTGG AGTTGATTA CCAAGAGCTG 480  
TGAAGCCCT CTACATCGAG GTTACAGGA TATACCATCT GCACTCTCA GCTTCCCTC 540  
CAACTACTAT GTATATGCT ATCAGACAT TAAAGAGCA TACGATGCG ACAAATAACA 600  
TGTTCGGCT GGCMAAGCA GTGCTGCC CTCTGCTGCT GCGCTCCCA TCGGAGTGT 660  
ATGAGATCC TGAAGTCCAC CCTCAAGTG AGGATTAAGT GCGCGAGCTG AATCCAAATG 720  
GACCTTGGG CTCTAGCAT GAGCGAAAC GTTGTGAGA GACCATGTGC TATGCTTACA 780  
TGAACAGGA AGCGTGGAA GTGCGAGTG CAGAGATCTT CACACCTTT GGGCGAGCA 840  
TGCACATGA CAGTGGGCA GTAGTCACA ACTTCATCT CCAAGCGCTC CAGGGGAGC 900  
CACTCAGCT ATACGATCC GGGTCTCAGA CAGAGGCTTT CCAATAGTC AGCGATCTAG 960  
TGAATGCGCT CTTGCTCTC ATGACAGCA AGTCAAGAG CCGGCTCAC CTGGGGAAC 1020  
CAGAGAGCA CACATCTTA GATTTGCTC AGTTATTTA AACTTGTGT GTTACCGGA 1080  
GTGAATTTA GTTTCTCTC GAGCGGAGG ATGACCGACA GAAAGAGAA CAGAGATCA 1140  
AAGAGGAAA CTTGATGCT GGTGCGAGC CCGTGTGCC GCTGAGGAA GTTTTAAACA 1200  
AAGCAATTA CTACTCGT AAGCACTCG AGTACAGGC AATATATAG TACATCCCA 1260

1320 AACCAAGCC TCCGAGATTA AGGAAAGAC GAGTCGCCA CAGTCGACT CCTCAGTTT  
1380 AGGACAGAG ACTACAGTAS TACCTGAT GAGATGATT TTGGCTTTT TTTTGTGTC  
1440 GTTAAAGAA AGACTTAAAC AGGTGTCAG AAGAAAGAC TGAATTTCA TTCTGAGCT  
1500 TCCATTAAAG AATGAGATG GCTTAAAGC TCCCTCGAA AAGATCGAA TTTGGCTTG  
1560 CAGTTTCA ATCTGCTT TTATGTAAA TACGTGAT GATCTCTGC GTATTTTCA  
1620 GTTTTATAT CTTCCTGCA GAGATATG TGTGACTTC GTTGAGATT TTATTTACTG  
1680 GTTCTTTGT GAACTGAAA AGGACATTA ACCCGAGCA AAAATGCCA TTTTATTTAT  
1740 AAAATGGGT ACTTAAATA TGAATGTTA TACTATGAT AAGAAAAAT CCGACAGTA  
1800 TTGTCAAGTG GTGTGAGCC GGCATTGAT TTAGCCGAA TAAAGAAAT CTGTGAGAA  
1860 GCTTATATG TCTCTTTAA TTGAGATTT TTCCAGTTC TACTTTTGA TTCCAACTT  
1920 GACTTGAATA TATCTGCTT GTTCATGATC AAGATATTT GAAATCACTA CTGTGTTTG  
1980 CTGCGATGCT GAGCGAGAG GAGTTGAGG GGCAGAAAT TAACTATTC TTGGTAAAC  
2040 AAGGTAAAT ATGCTATTTT AATAAAAAT TGAATCTAC GAAAAAATA AAAAAAATA  
2041 A

(2) INFORMATION FOR SEQ ID NO: 132:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2012 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

60 TACCAAGTG CAAGATCTA CTATATCATG GAGAGAGAG TGAATGAGAA CTATGCTCT  
120 GACCGAGCT GGGAAAGGAA ATGCAAGAC CAGTCTGAA AGGACATTA TGTATCAAT  
180 TTCTGAGAA AGAAGAGAG GCTCTGAGT TCCAGATACA AGAAAGCTG ATTGAGGAA  
240 TACATGAGG GTACATTCAG GATCCCTGAG CGAAGAGTAC GACCAAGAA ACACTTGGA  
300 ATCTTGGTC CACTATACA AGGTAAATG GTGTATATC TGAATGATG ATTCAAGAT  
360 AATGCAAGC GCGCTTACTC TGTGATGCT CATGAGTGC TGAATCTAC TACTGTCTG  
420 CCACTGGTG CTGAGCTGAG TGAATGATC ACTTATCAT GAAAGATCC AAGAGAGCT  
480 GCGCTGAGC CAATGAGCT GCTTGTGTT CTGTAGCTA TTATCTGCA GTGATTCGA  
540 TCAAGAGAT GTATATGAG CTGTGAGGC CTTGAGTAT CTGCAAGAG GCGATCTG

600 NAGCCCATG GAGAGAGAG TGAATGAT GAGAAATTTG CATGTGTT CTGATTTT  
660 GATGAAAAA AGTGTGTTA TTGAGAGAA AATGTGAAA CCAATGGTTC GAAATGCA  
720 GCGATTTTA ACTTACAGAA TGAATCTTC TTGAGAGCA ATAAATGCA TGCATCAT  
780 GGGAAAGCT ATGCAAGCT TGAAGCTT ACATGTATC AAGAGAGAG AGTGGCTG  
840 TACATGCTG CAGTGGGCA AGTGTGAT CTACAGCA TCCACTTCA TCCAGAGAC  
900 TTCTCTATC GAAATGAGAA GAACTAGCG GCAAGTGG TGAATGCTT CCGAGAGCT  
960 TTGAGAGTG TGAAGATGT GGCAGAGAC CTGTGACAT GCTTATGCA CTGCAATG  
1020 ACTGACATG TGCATGCTG CATGAGAGC CTCTCAGTC TTTTCTCTG AACAGAGAC  
1080 TTAAAGCTC TCAAGCTAT CACGAAAGAG ACTGAAAAA CAGTGGCCC GAGAGAGAT  
1140 GAAAGAGCA ATGTGAAGAT GTGTGAGTG GATATGCCA TAAAGAAAT TGAATGCTG  
1200 GCGCTGTTT TGTGTGCAAT TATGTCTAC CTTCCTCTG TTGTCTGAC ACTGTGAG  
1260 GTGTGTTGT ACCAGATCA AAGAGAGAG CTAGAGAGCA ATAGAGATC CATCTGAT  
1320 GAGAGCTCA AGCTCTGTC TTTCAGAGC TAACTATGAG AGCTGAGAA TATCTGAGG  
1380 AAGCAATCT GTATGAGCT CCGAGAGCG CATGAGTAC TCACTAAGC CAGACTGAA  
1440 GCGCATGAG TGTGAGAAA GAGAGAGAG CAATCAAGT TATGTGATA TTCTTTCT  
1500 TATTAATTT ACATGAAAT AATGAAAT CAGTTTCT TTATTTCTT TGTCTAGAT  
1560 GCGCACTGAG CACTAGAGAA GTACTTTAT ATCTATATC GGAATTTCA ACACTGAT  
1620 TATATTTCT TCTGAGCTT GAGAGATTT GAAATTTCA GAAATGATC CTCTGACA  
1680 AATGAGAGC AAGAGAAAA CTATGATTT GGTGTTTAC TTCTTTGAG GACTGAGAA  
1740 ATTCACTTT GAACTGAGC CAATGAGCT GTTAAATTA CCGACATTA AACTAAAGC  
1800 TAAAGATTA GCTGTATG GAAATGAG GTAGCTGAG TATGTGAAT CGAAATGAA  
1860 TTGTATCT CTGTGAGT GAACTCTT GAGAGAGTG TGAATGCTT GTTGTGCCA  
1920 TGAAGATTA CAATCTGAG AGCTAGAGC TCTGAGAA ACCAGATCT CGAATCTT  
1980 AACCTGAGC ACTGAAAGAA ATGTGAATG ATCTTCAAT GTTTAGACA GGAAGATTA  
2012 TCCATTAAAG TACTGTTAG AACATGAA AA

(2) INFORMATION FOR SEQ ID NO: 133:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1669 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

5 GAGCAGATATT TTAACCAACT TGTATTACAGT ATGTTACAGT TCATGTTAGG AAGTCAGAAA 60  
AGACTTTGTT TGTCTTTGTT CTGCTGATGT GAGTCATGTT TTGTGGGGTC TTGATAGGCA 120  
CAATTACCTG TTGCTGGGTC CAGATGTGCA GGGCCAGTCT AGGCTGACAC ATCTACCCGG 180  
10 AGGAGAGGCC TGTTCCTCAT TTCTTCACTC TCCCTCCGCC ATATAGCAAC TCTCCAGGTT 240  
TTAGATTACC GTTTTCGACG ACGATTACAC CAAAAATGCC CCAACAGGTT TTATTACTGT 300  
TTATATACTA TACTTTTATAC AGTACAGACC CTAAATTTTA TTATTGTTTG CTCCGCCAAT 360  
CTGATACCAA ATGTTTAAAG TTGTTTGAAA TCCAAACATG GTAGTGTTCA TGGGTAAATA 420  
TTTTCAGGCC TATGTAAAGAG TTAGAGAGCC ATAGCATAGA AGTAATCAG TACGATCTGA 480  
20 GACTGTGCGA GGCAGTAGGG CCTCTCTGGG CCTACAGACC TCAGTTCGCC AGCCTCACT 540  
TGCCTGCTCTC TCACACTGCC ATCAGGGCTG TTAGTGGCAC CTGTATGAGG CCAAGTGTGC 600  
GTCCAGGGGA ACGACACAGG TTAATGGTTC TCCCTAGAAC TCATCAAGTC AGTTTAAATC 660  
ATCCATGAC ATGAGTTTCA TTATGTGTTT ATATAGCTTT CTTAGACATA CCAACACATC 720  
ATTCTAAAT CAGATAAATT ATTCAGTTT TGTGTTTAGA AGCTAAGTA TGTGTAGCTG 780  
30 GAAACAAAA TGAAGTGT TTCTCTGCTG TTAATCTAGA GTGTGCAATT ACHATGTGT 840  
GGATTAATTC ATGTTCCAGG GGCCTCTGGC ATCTCCCATG GACTGATTC CAGGAGAAAA 900  
AGCCCAAGG GAAACCCAGG ATTCCTTTG AGTAGATGTG GGAAGAGGCC CATGAGAGA 960  
35 TATGAGTCC TGTAAATATC AGTGTGTGT GTGGCTGCTT GTTAGCACTC ATGTTCATAT 1020  
GGTGTTAGGA GCTTCCCTCAT CAGCCCTTTA CATGATGTAG GGAACAGTGT CTGTGAGAT 1080  
40 TAACTTTGGG ACACAGTGGG TTAGCCCTGA GAAATAGGA GGCCTGCTCT GGACCTAGGG 1140  
AGAGAGCCA GTGACACAGG CAGAGGGGTG CAGCCTCTCT TCCCTTCCAT TTGGAGGGGG 1200  
45 TGTGTCAGG AGCCTGCCCC CTAACTCTG CTGAGCATTA AGTGACATTT GCTTTTGGGG 1260  
CTTATCTCTG ATACATGCTG GAGCCCTGCC TCTCCACTGC TAGATGGAAC CTGGATCTTC 1320  
TCATCTACT CTTAGTCTGT CAGTTTCTAC GTGTGAGAG CAACTTGTG GGCAGTGTTC 1380  
50 CTGTGACGTG CTGTAGCACT TAAAAATATA TTCCAGGGTT CCCCAGAAA CAGTCCGAG 1440  
GGTTCCTATG ATCTGTAGTT TCTAAGTGA TTATTAAGTGG TTTTGGGTAC CTGAATTTTG 1500  
ATGCGTAGC CTTAATATA GTCTGGGGT ATCATGTAGA ATCTTTTCTG GTGACACAGT 1560  
55 CATAAAGTTC TATGAGGAG TTCTATCAAG GCATCCAGTT CAGTGTGCT ATGCTGGTTA 1620  
CAACTGAGA TTTTCAAT AAAAAATTG TCTTAAAAA AAAAAAAA 1669

(2) INFORMATION FOR SEQ ID NO: 134:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1565 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

5 CACTTTTCT ATATCACTTA AGTCATACCC CTCCTTTAGT TACCTGCCAA ACTCTGCT 60  
15 TGGTTTATAT TGCAGTTAAC ACGTTTACAA AGCTGTATAG GTGTCTTTTT TTCTTTGTGA 120  
ACGGAATGTG TAAATCAAGG TATATACATT GTGTGATGTT CTGTGTTCTG CAGTTTCATG 180  
20 AGAATTACAA CATGCCATTC AGTGTCTGT ATAGATCTGC CTACCTTTGT GAATTCATCT 240  
GTTAACCCCT CTTCCTTTGA GAGAGCACCG GCGATGTGTG TTACTCTCTT GTTCTTTCTC 300  
TCTCTCTTAC TGGTTATCTT TGAATTACG ACAGACTCTT CAGCTGGTTT GCTTTATCAT 360  
25 GAATTAATGT TGTGACCTTG CAGTCTCTTC ACGTTTACAG AAACAGTGC TACCTTCACT 420  
GACCAAAAT TAGGAGGGA AAACAGCTT TTAAACGA TCATCTTTT AACAGCGAA 480  
30 ACCGATGTGT CTATGTGCT GCACCTGCT GTTCTACTTC TCAATTCAGA CGTGTGTGAA 540  
CGATCATTTT TCACTTAACC GTGAGATGCT CACGATGACC CTTCCTGTTG TTTCTTAGCC 600  
ATTEAAATCG AGACTATTTA TTGTGAATAT ATACACAGT GTTTTCCAC TGTATTTTAT 660  
35 TTCCAAAAGT TGAAGACTGC TTCTCTTACC TTTTGCAAA TAATGTATAT TCCATTTGAG 720  
ATTCTCAAG ACTTGATAT GTTAAACCTA TTAAACCTAG AAATGTATAT CATCCTTTCA 780  
840 TCACTGTGGC CTGAGTCCC CAGCCCTCT CTCTCTTTT TTTAGATGAG ATTTAGCACA  
900 CTCTCAATTA TTTAAACATG CACATTTCT TCACTATGTA TGTGTAGGCC ATCTGAGCTC  
ATAGCTGATT CAGTAACGAG TTTCATGCTG TGTCTATCAC ACTCACTACT TAATATCTGC 960  
45 ATGCTGAAA TGTGAGGGA AAATGTATCC ATGTGTGTCT GCGAAGCAT TACACTTTGA 1020  
CATTTTAAA TACTGTGATT CTGTACATAT TCTGAGTTT GTTTGTTTTT ACAGAAAAA 1080  
1140 AAAAAAGT CATTAAGCAA TCGAAGAGCC AAGAGGTTTA CTATGATGC TTAGGGGTCT  
1200 CTGACCTTGG CTGGCATA GACCTACAG GCCAATTTAA TTTAGGAGAG TAATATTTT  
1260 TCAAAAGCCA ATTTTTTTC TGTATTTCT GTATGAACCT GCCATATCA TGAATAGAAA  
1320 GCGAGAACCA TAAAGAGAA AAGAGTGAT GTTCTGTAT GTTCATGTAA ACCTAAGAA  
55 ACAGTGTGGA GGCAGGGGCG ATCAGCGGAA CTCTAGGAGC TTGCTGTGTC TTGAGAGGCA 1380  
TCCATACCTG CATTTGCAI TCTTGTATG TATCTATAT GCCAAGAGCA AACTATTTCA 1440

TCATTATTG TAAATACAC TTTTCCGAG ACCTACGTA AAGTTTCGT GATGTATTG 1500  
CTTCAGTTG GATTAATAAT TACTAGATTG CATCAATTA AAAAAAAAA 1560  
CTCA 1565

10 (2) INFORMATION FOR SEQ ID NO: 135:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2007 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

20 TCTAAGACC CCGTTTACC CCGCTTTG GAGGAAAGT CCCCCTGAG GTCCAGCT 60  
GATTGTGAC CAGCTGAC AAATTCCTA GCGACACT GCGTCAGTT CAGATTGCA 120  
GCTGTGTG GATTGAGAC AGGAAAGG AAATGTCAG CAGACACA GAGCTGAG 180  
AGGACTTGA GCTAGTACC CTGTGAGAC CCGCAGCTT TGTCAAGCA GGTGTGAG 240  
CTTCCGCAA GTCCAGAGT ATTGATTG GGGAGACTG AATATTACA CTTAACTGT 300  
GAGCAATCC TGAATTTAC TTTCTTAG CTTCCTCCC AAGTTCTTC TGTCAAGAC 360  
ACACAGACC TTGCTCAGA ATCAAGAC ACCCTCAGG CTCAGAGT GGAAGACT 420  
GATTTGAGC TCCCTTTG TTGAGCTAG GAACTTCGT GCTTTGTG GCTCCACTG 480  
CAGAGCACT GGAAGTGTG GATCCGATT GCGCTGAG CAGCTGAG ACCGACAG 540  
GCGGCGCTC TGAACACT GTTAAACA CACCTCTTC TTGCTCTCT GCGATCTCT 600  
TGTGAAATA GTCTCTGCC TGGCAGTTG AATGGGAAA GCTGCTGCA CAGAAAGAG 660  
AAGCAATCC GCGTAGACT TGAAGATTG CTGAGCCG CTCAGAGG ATTAATCACT 720  
GAGGAGTTT TCAAGTCA ACATCAATCT GCGTTATTG GGGGCGAGT GTTCACCA 780  
AAGATTGCA AATCAAAAC CATTGCGGC TGTCTCTCT CTTCACAG CTCGCGGAA 840  
AAGAACTCC CTCCTCTCT ACTGATTCG AAGTGTGTT GAAATGTG GAGCACTGG 900  
ACTTTTTC TCTTTCTCT GATGACACA CAGTCAAAAT GCAATGCC CATTAACTT 960  
TGAAGTAA TTTCTTCTT GATCAAGAT GTTGTGCC CTTTAAAG GAAAGAAAT 1020  
AAGCTAAG TGTCCAGG ACTGTGTAG GCGTTTAT ATGAATCTC GTTAGAGTA 1080  
GACTAAGGA TGAAGATCT TCTTAAATA AGGCGCTTAA GTAAAGATA AACGAAACA 1140  
GTAAAGATG AAGAGTGTG TGTCAAGAT CAAAGTAG AATGCTTCA GTCTGATTT 1200  
TGTTCAGAG GACTTAATC GTAAATATT CTCCTATC ATACTCTCC TTCAACCCA 1260

CGATGTCT CCGTCACT CTCAGATGT CAGAGGGTA ACCAGTCC TTGAGAAAT 1320  
TGGAGACA TGAATATGT GATGTGTAA TTTCTTTAA AAACTTAG GAGCTTTGC 1380  
TACCTTCCG TGTGAGTT GTTTGAGT TCAATTAAG AGCGACATC TCACTATTTA 1440  
TACAGGTT GCGCTGTGT TGTCCATG TGTATACA TTTCAGAGC TCCCTGTCT 1500  
CTTGAGCTT TTAAGACA ACCAGTAT CCGACTTGA ACTGCACT TCTTATGCT 1560  
TCTAGTATT TGGCAGCA TCAACAGG GATTAAAT ATCAATCT TCAAGCATG 1620  
TTCAAGAGG GCTGATCA ATTGTAGAG GTTATGGA AAGGAGGG GAGAGAAAT 1680  
TCACTTAT TTATTAATT ATTAAATG TTACATCTT CTTAATGTG TATCAAGCT 1740  
GATTAAGAC TGAAGACT AAATTAATC GTTCTCTCT CTCCTTCC TTTCTTTT 1800  
TTTTTTTAA AATTAGAT AACATTTT TTTTCTAAA GATATTTG AATTGTGTC 1860  
TAAAGCTG ATAAAGGT CTGTTTAA AGGTGATTT TCAATCTCT GGGACAGTG 1920  
GTCCAGAA CATCAATTT GTAAAGAC ACAGTAGAG ATCTTCTCT GATTTCAAA 1980  
AATTATTC TGTATGAT TAAAGT 2007

30 (2) INFORMATION FOR SEQ ID NO: 136:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1291 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

40 CTTTAAACC TCCCCCTCA CACATACA TATCAGTTG TTTCTAGTT AAAAACCAA 60  
GTAGCTGAG TTCACTTGA ATTCAGTCC AGATTGCA TGAATCAGC CATTAATTT 120  
TTTCTATT TTAGCTGTT GCGCTTAGT TTTAAATG AATTAAGAA CTCACAAAG 180  
GTTTATTT CTACATCAC TTAGGTTTA GGAACCTTA CCAAGTTTA TCAATTATC 240  
AATCAATG GTTACTGAA ACAAATAG TAACTTTCA TTAGGATT ATTAAAGTT 300  
AATGATGTT TTTCAGAAA ACCTTCTC ACATTTGAC TTCCAAATA AATCAATGTA 360  
TATTAAGTT ATTCAGTGA AAGCAAAAT GTTGTTTCA AAATTAACT AGTTTCTGT 420  
ACATTAAT TTGAGAGT GACACTGAC TCTTTCCAG TCTTCTTCA TGTCAAGTTT 480  
CTATAGACC ACTATTGCA AACGATCT GTCACTTACC AATGTGTAA AATTTCTGT 540  
AATCACTTT GTCTATTG TAAATAGTA ACTAAACTT TTGCGAATC AGCAACATTT 600

5 GCTGAGCCTG TTTTTHAAG TAAATGTAT TCTTACTAAT GTTCTATCA ACAATGAA 660  
TCTAATATAT GCTGTCTATT TCTAATGTTC ACATTCATAT TTTCAGGTTC TATCTTATTT 720  
780 TAAATAGAA CAGACTCTC AAAAAATCTT CAGAACGAG TTATTATGA AATATCGAA  
840 TATTGAATA ACCCGGTGG GTTAGATATC TCACTGTGTC ACCAGTGGG ACATTTGCAT  
10 GGACTGGGG CTTAAGGAG TTAGAAGGA CTGTATAGTA TATCTGTGAA ATGAGCCAT 900  
960 CCCCACTTGA ATGTATAGT GATTAACCC ACTTTACCA CCCAATTAC ACCACCGAG  
1020 GCCCATTAAC CACTTGGCT CTGGTCAAT TTTCTTTTCT TCAATTTGTA TGTCTAGAT 1080  
1140 CAAATGTGT GTTCTACAT GTTACAGCT TCTCTTTTGT TGTATTAAG ATTTAGTCC 1080  
1140 TACTTTTGT TGGACATAT AGATATATCA GAGACCAAAA TAGAAGAT TGTGTGTGA 1140  
1200 TATTTTTCAG AAGTCAGGAG ATTGTGGCA AATCATTTAT TTGCTTTT AAAAAATCAT 1200  
1260 TTAAGCAGTT CAGAGAGTAG ACTACTCAGA AATTAATTTT AATTAATGT CTAGAGGTTC 1260  
1291 AATATTTTTT AATCATATT GAATCAATA A

25

## (2) INFORMATION FOR SEQ ID NO: 137:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1906 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

55 GGCAGAGGA CCTACTTTTG TACAGACCA TGGTGTGTC CAGGTAAA CACAGTGAT 60  
40 ATTTTGGAT GCTTTGTCTG CATCTTGAC TTGTTTTC AGTATCATTA TTCAGACTTC 120  
AATTTGTGAA TCTTTTAAAC ATCTTGATA AGATCATCA TTTGTGTG AGAGCTGTTC ATCTTAATAT 180  
GTATGAAT TCACTAGT TCTGCTGATA AGATCATCA GTTTTGAG GTTACTGAT 240  
45 TTGCTTTC CTCTTGAT TTATCCCAT ATATGAGAA GAGTATGT CAATCTAAC 300  
ATTTGTTTT AATGTTTAA TAAAGTGTCT GGGCAGTGT GTACATTC TACTATGT 360  
50 CATTAAGCA AATACTTAC ATAGCTTCT TAAATATAG GAATGATAT ACATTTTAT 420  
GAGAAAGTAA GTTGTGTTTC AGCGCTACT TAAITTCCTT CCATATATG TGAATCAAC 480  
55 TTTTGATAT GGAATCTTAC TATTGATA GAAATGTGA TGTATATAT ACATCATAT 540  
ATAAGCATAT ATGTGTGTGT GTGTGTGTAT AATATATAT ATGATGTG TGAATCTTA 600  
CTACACAACA TAAATCACTT TTTAATTTCC AGGAACGGT AGTGTGACAC GTGTATATC 660  
60 GTTTTGAGGC TGAATCGTT ATTAATGTGT TATTTAGTT TTAATCCAG TAGCAAGGA 720

5 TTTTAAGTAA GTTGCATTA CATCATTAT GTGATTTAAA ACTAAGAATA AAGCTGCAT 780  
TTTCAAGAT AATTCGAT TCTGTGTGT GAATTAACA CCAATATCT GAATCTGATG 840  
TACATACAGG TTTTACAGG AAGAGATGT ATATTTTACA ATTTCGAGT TTAATACCA 900  
GGCTACCCA GAAAAATGA CTGTATACA TGTACCAAT AAGTAAGGA TGTCTCTG 960  
10 GTTTCTTTT CCACATTTCA AGATTTTAA TTTCTAGTT ATTATATCA ATTTGTAT 1020  
AAGTAAGCA ATGAATTTT TAGGTTAAA CACAGATGG GGGTTTGTG GATGTTTTAA 1080  
1140 TGTATGGG ATTTTATGA GATTAAGCC TTGTTCTGC ATTTGAATAT TTTCTATAT 1140  
1200 TTTGTTTAC AGTTAATCTT CCTCCCCCA GTTTGCTATT CAAATCACT GCCTGAATGA  
1260 CATTCTAGT AGTCTGATGT ATTTTCTGA GGAATAGTT GTGATCCAA TGCAGTGTTC  
1320 TTTATACA TTACCTCTAC ACTGCAAG AAGCAAACT CTTTATATG AATTAATCA 1320  
CATGTATG GCGAAATAG TTTGAAAG CTAGATAT ACAGTGGC AATGTTGTGT 1380  
CAGCTTGGT ATGAGTGT GGCATTAAT TCTAATACAT CCMAAGACC ATGACTGAA 1440  
25 CCTAACTCC CTGGAATC TGGACAAAG GAATATGAA ATTCATTTT GAAATCTGAC 1500  
CAGTATAT GGCCTCAGA CATGATCAG CAGTGGCC AAGCCATTT CAGTACAGA 1560  
30 AATTAAGAG ACTACAGTA AATAAATTTG AACATTAAT AATAATTTAC CACTTTTGT 1620  
CTTTAAGC ATATTTGTA ACTAGAACT GAGCAAGT GACTTTACTT TCTCAATTT 1680  
GATCTAGT TCACTGTGCT CTTTCCCTC ACCCTTCCC TTCCCTTCC TAAAGCAATA 1740  
35 GTCCAACT TAGTATTT TTGCTTCCA ATTTGAATGA AACTTATAT GCAATGAT 1800  
TTTTTCTTT GCAAGACAC TGTATATCAT CTGTTTAAA TGTAAATGTC CCTTATGCT 1860  
40 TTTGAATTA ATTTCTTTT GTAAAAAAA AAAAAAAAAA AAAAAA 1906

## (2) INFORMATION FOR SEQ ID NO: 138:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1935 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

55 TCTGAATTA TGTATACGA TCCCTCTGAG GATTTCTTGA TGGCTGAGC AGCTGGCTG 60  
AGCTAGTACT GACTGACAT CATTTGATG AGGCAAGCTT TCTGTACAG GATTTACGC 120  
TCTATGTTT ATATACATTT TCACTGTAC TTGACCTCA CTTTACACAA GAGGAATTA 180

TCGAAAGTGA GCTGATGCC TGAAGTGC TTAGTAAAT TGGCAATGCC ATGCTTCCA 240  
CTGAGCTCTT GAGGTACAGA AGTCTACTT TGTGCTATT TGTGATCTT TCTTAAAT 300  
GTGCTCAACT TTGGAAGTCT TAGAATGGT CCGTGTGCT TTTTACTTT GAGAAATCA 360  
GTTTCTGCTT CTTTTGGAA AAGAAAACA AGTGCATTG TTTTCTACT GAAAGTACC 420  
CAATGACAT AGGTGAACAG GAGCTAGTT AGCGCTTCT GTAAACAGA AATCATACA 480  
AAACATATC TTGCAATCT TTTCATATG CTTGCTACTT CTGTGATA TTTCATTCA 540  
GAGAGCAC AGTTAAACCC GTGAAATTTG TAGTAAAGAA CTTGAGTCA AACCTCTTC 600  
CACTAAATGG CTATGCTCT GAGCAAGTTT TTTTTTTTT TTTTTTTTA ACCCTTCTG 660  
AGCTTCACT TTCTAATGCT ACCTCAGAA ATTGTGTGA GCGTGAAGT AATGCATTG 720  
TAAAGGCTT GCGAATAGG AAGATCTAG TTATGAAATT ACAGCTTGT TAAAGCTGA 780  
AGAGCTTAA ACCTACATG AATGCAATG CATTTACCC CACTACATG GACATATG 840  
AAAACATACC AAGATCTCT TTTTCAATT ACTTACAGT TATTCATAT AAAATTTTG 900  
TAAAGATAA TCTTAAATT CTAACTTAA GCTTCTGTT TATACACT CCGTAAATC 960  
TGGATATAGA TAAATGACA GATTACTTA ATTCTAGT GGTGCTGCT GATGTATAT 1020  
TGTAGATAG GACATTTCT TTTTCTAC AGCTGTAG GTCAAGAAC TGTGGAAG 1080  
GAGGAGGATA GATGACCC ACAGAGGAC TGAAGTGA ACTGAGT TCTTTTACT 1140  
GTGAGGATAT GAGCTGCTT TGTGTATG GTTTCCTAG GAAATGCT GTTCCCTTC 1200  
TTGCTATGAG CAGTACATA CAAGTGGCC AACCCAGTA GCGTAACTT ATATATATC 1260  
AGTGTGATG CTGACTCTA ATAGCCCAAC CAAAGTACC TAAAGTTTA CAGATACAT 1320  
AATTAAGCA CCTAAATAT TGAATGCT GTTGTGTGA CATATCTA TGGCAGAAC 1380  
TGAAGCTAG AGTTATTAAT CATGTATAG GTTCTCAG AAGGACAGA TTAGTAAAT 1440  
ATATGTAT ATGAAAGGA GTTATTTAG GAGAACTAC TCCCAAGT AAGAGGGA 1500  
GTGCAAGAT AGCGCTTG CAGCTGCTT TAAAGAGAG CAGATATG CTGAGCTGA 1560  
GTTCAGAAAC CTGAAACTG GAGAGCTGA CAGTCAAC ACCCTCAAT CTGTGACAA 1620  
AAGCAAGAG CCGCTGACA CAAACCACT GGTGCAATC TGAATCTCA AAGCTAGAG 1680  
AAGCTGAAT CTGATGTCA AAGAGAGAA GATGTAGAA AAGCAAGAG ACTGAGAA 1740  
CAAGTAAAC AGTGTAAAC AACAAATG CAAATCAAA GAGCTTACC ATTCTTCT 1800  
GCTTACTGTA TCCCTGAAT TCCCTGATC TGTCACTT CTAACTAG TTCTTAAAG 1860  
CTTTCATTA CATGACTCT CTGAAACCC TCCAAATAT TGTCAATGA AGTTCAAAA 1920  
AAAAAAAA AAAA 1935

(2) INFORMATION FOR SEQ ID NO: 139:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1446 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(+1) SEQUENCE DESCRIPTION: SEQ ID NO: 139:  
NCCCCCTG GACAACTCA GATGAACAC GTTCTGCCA GAGAGCCCTC AATCTCAGA 60  
GAGACAGAC AGCAATTTCC TCGTGGGAGA GAGAGAGATC CAGCAATCTT GATGCTGCT 120  
GCAACTAAT TTTCCTGAG CAGAAATCA TTTCCTGAG TGAAGACAG TTCTGCAATG 180  
TGAATGTG TTTCCTCAG CAGAGGCCC CTCTTTTCC AGCATTTCC TCCCTGCTCC 240  
AAGCTCAG CAGACACCA GTTCTGCTC ACATGCAAG TGAACAGAA CTTCATATTTG 300  
GAGAGACTG AAGAGGTTGA ACAAACCTG AAGGAGGCCC GAGCTTCTCT CCGAATTTG 360  
GAGAGAGGAG TGTGACAG TGCCTCCCGC AAGAGCCAG CATGTTTAC CAAGCTCTC 420  
ATTGTGCT GAGACAGACC TTTTCCCAAG CCGTCAAGAC AATGCTATC GGTGCAAAAC 480  
TGGATGATG GATTTAGAG GAAAGATCC CAATATGAC CAGAAATTTT CCGTCTCAG 540  
GAGAGCCAG GAAACTAG GCAAGCAG ATACAGGCG AAGGATGTG CAGTGAAGG 600  
GCTTCCCGCC TGTGCTCTT CTCTCACA TGTCTCCCC ACCCTGCTC AGTTCTGCT 660  
TCCCTCTAAT CTGCTTCCC CTCTTTAG CTGTCCCAT CTGAGTTGA GACAGCTTT 720  
CTCTTCACT GACAGCTTC CTCTGACCA GCGCCCTTC TTGTCTGAAA AAGGAGCTT 780  
TGAATGTAG AAGGAGCAG TGGGAGAAA GTTCTCAG GACAGTTG GAAATGAG 840  
TGAAGGTGAC TGGAGAGAG ATGAGAGGAG CATGTGAGAC AAGCTTTGAG CAGACACAG 900  
CAAGAAATAT TTGAATGTG TGAAGTACT CCGGAGAGC CTGAGCTTTG GCAATTTGAG 960  
AAGAAATGA TGTGTAGAG GACTTAAAG AAGCAGTGA CAGAGGAGAA GTCTCAATCT 1020  
GCTGCAATTT TGTGAGTGT TATGTCAAA CTGAAATAG GACTGAGTGT CTGTCTCA 1080  
CTGACACCA AATTCAGAT CCGTGTCTT GATGTCCAG AACTTTGCT CTGAGCTTC 1140  
CGTCTCTTC CTACCTTCAAT CCAATGAAA TTAAATTAAT TCTGATCTTT TCCCTGCT 1200  
GATGATCTC CTCTCAAC ACCGATGCC TCCAAATCT AAGAGCTGAG GCGCTCAAC 1260  
CTGTAACAG AATGCTCAG AATGAGGACA TGGAGAGGAG GTGAGGAGAC CCGATATTC 1320  
AAGCAAGAC TGCATATGCC AGCTGCTTC CCAAAACAA TCCCAAGAT CCGTTATTC 1380  
TACCCCAAC CTTGGGCTT CTGTACAGAT TTTTAACTT GCAAAAGAT GAGAGATA 1440

TTGTAA

1446

5

(2) INFORMATION FOR SEQ ID NO: 140:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1109 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TTTTTTTTTT TTGATATGA AATTGTCTTT CTCATTGCA GAATTAAGCT AGGGAACAC 60  
TAACCCAAA ACTTCTGTA GAGCTGTTC TTGAGAGCA GCATCACTTA TTGGAGTAA 120  
AGACTCGTA TAAGACACC ACCATCCCTA CTTGGGTGAT GGGGTAAT TTTATAGAT 180  
TCCATTTTCC TAGTCCACA TGTGAATG GATTTTGATG ATCTTAATCT ATATCTACC 240  
CTTATATAA AGATCAAAA GATTAATCT CTATGAACAG ATTGGAGATA GGAATGAAA 300  
AGTTGGAGG ATGCTTTTAT TCTAATGTGA GGTAGGGAA AATGTGATA ACNTACTGG 360  
GCTGAGGAG GCATGTCTT TTAGTTGAG TTCTCATTTT TATCTCCAG TACTGACTTG 420  
TGGGGAACG ATACTTTTTC ACTGCCAGT ACTGAATGCA GAGGCTCAGT GAATATATA 480  
TGTGGAGAT GCATGCAATT GGTATATAG CAACATAGC TGGATTAGA CAAGTTGTT 540  
GTTTTGAAA GGGTTAAG CTTAAGTGA ACAATCTAG CTACAGTGA ATGAATAGG 600  
TAATATACT TGCATATTT TAATTTCCTT TGGTTAAGG TCCCCATAC TTCTCTGTT 660  
GGAGACATGA GAATATGAT TACTTCAGTG TTAGTTTCT TATTTTTTT TTTCCTCAT 720  
TTGTCCCTTG TCATCTTTT GCAGCTAGA AATCTGTGG TTATACATAG GCGACTCTT 780  
TGTGAAGTG GTTATTTCCA CTGAGAAAG GGGATTGAAA ATCAGTTAGA ACCAATGAT 840  
TTCTTGCCC ACGGAACAT ATTCTATDA GATAGCTGA AAGAGTCTG GTGAGAGCT 900  
CAGCTCMAA CACAGATCA GCACCTCTA TAGGAATCC CATGATAT GACTCTCAT 960  
TCTGTTTTAT CAGATGCAAT ATATGCTTA CTTGAGAAA ACTAAACAG TCATTTACA 1020  
AAGAAAGTCA ATCTGTATCC TAGCATTTT AATAAAGT TAAACAAA AATTAAGG 1080  
GACTCTGAG,GGGGGGCCCG AACCCAT 1109

55

(2) INFORMATION FOR SEQ ID NO: 141:

(1) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 497 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

TAGACTAAC TTAAATCTT TTATCATCT TTATTTAT AAAAAATTT ATTCTTGA 60  
ATTTCCTGT ATTTCCCTTA RCTCTCTA TAAATGTTA TATCAATG ACCATFAC 120  
CATATCCTT AACATTAAT CTCAAAAGC TTTTATTTT TATTTTITG AAGGTAGTTT 180  
TTCTGTGT ACTCTGTA ACATTTTCC TTTCAAATCA TTGTGTGCC CCATACAAA 240  
ATGCTTTTTA TTTTGTAGA TGTGACTT TTTAGTATG CATGAGTGTG CTAAAGCCA 300  
GATATCTTTC CACATCACT GTTGCTTGG ACACCTAGT TTTATCTCC CATCTTACT 360  
TTAAACCTG ACATGCAAT CCTCAGTCAG GGCAGAGAC GGCCTGAGC OCTTTGTGA 420  
GATGCTCAC CACAGCAGA AGGCTGAGAC CTGGTACTT GTACCTGTTT ACTTGTATA 480  
AAGAGATTA TCTTAAA 497

(2) INFORMATION FOR SEQ ID NO: 142:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 269 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

ATGAGGAGA GCGAGCTGC CTCGACGCC CTCCTCTCA GGAATGGCTT TCCGAGAA 60  
TCCGACAC ACATACCTC TCTTTT TTTT CTAGTCAAC TCTGTATT TCTTGGCTT 120  
GCCTCCCTC TTCTCCCTC TCTCAACTT TTACTTCTG TTCTATTTT ATGGGATTTG 180  
GGGTGAGT TAACTTACA ACAGTGGCC CACAGCCAG TCTTGCAGA AAAAAATCA 240  
AAGATTTA ACAAAAAA AAAAAAA 269

50

(2) INFORMATION FOR SEQ ID NO: 143:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1269 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

60

TTGATGACT ATGATCTCTC CCGATACAG GAAAGTCTG CCGAATGAA GCGATAGAC 60  
TTTATACACT CCAAGACAT CTGCCCCCTG CTGTGAGCT GCTGAGGCTT CTTCGCTCTC 120  
TTCCGACTGC TGCATGAGGT GCGCGGAGAG GCGTACCTGC GAAATGCTGT GGTATGATTC 180  
AAGCGGCCA CCTGAGGGCT GAGCAAAAGA TGTGAAAG TCTTCTATGC TCGGGGTGCT 240  
AAATGCTGTC TCTGTGAGCG GAAATGTGCG GCGCTAGAG AGCTATCAG AAGATTCAC 300  
GCTTCATG CCGACAAAGT GCGAGACAC AAGCTTACT TGTATAGCTT GAGCTTCACA 360  
GACTGAGGG CCAATGTTGC AGCAAGACT GAGATCTGTC AGTCTTTG GATGTTCAC 420  
ATACTGTCA ACAAATCTG GATCAGCTAC CTGTGTACA TCAATGACAC CACATGAT 480  
GTGACACAGA GGTTCATGCA GACAACTAC TTTGAGCCAG TTGCTGTAC GAAAGCATC 540  
CTGCGCTCCA TGCATAGAG GAGCGAAGG CACATTTGCG CAAATGCGAG CATCGAGGG 600  
AAGATGACA TTGCTTTTTC ATCGACATAT GCAAGCTTCA AGCAGCGAC CAGAGCTTTC 660  
TTTGACTGTC TCGTGCCTCA GATGGAACAG TATGAAATG AGGTAGCCTT CATCGAGCCC 720  
GCGTACATCC AAGACAACT CTCTGTAAAT GCGATCAGCG GGAATGATTC TACGTATGCA 780  
GTATATGACA CCGACAGAC CCGAGGCCGA AGCGCTTGTG AGGTGAGCCA GAAATGCTT 840  
GCTGCTGTG GGAAGAGAGA GAAAGATGTC ATCTGCTGTG ACTTACTGCC TTGCTTGCCT 900  
GTATATCTTC GACTGTGAC TCGTGGGCTC TTCTTCAAGC TCAATCTGCC AAGGCGAGAA 960  
AAGAGCGAA ATCCAGAAC TCTTATGACT CTATACAGCC AAGGCGAGAG CAGAGTACA 1020  
GCACTCTTAC GCTTGCCTTAC TCTTACAGGG ACGATTCAT TTGTATGAC TTATATGAG 1080  
ATTGTCTCA GATGTGGA AAGCTGAGA AAGCATCTC GTTACATCT GCTGCGAG 1140  
GACATCGAA AAGCAGACA AGCTTCTCC CAGGTGTAGG GAAACACTT AAGCAATAA 1200  
TATGAGCTG GGTTTTACA CTAAAACTA GAATTAACA TCTCAACAG TAAAAAATA 1260  
AAAAAAAC 1269

(2) INFORMATION FOR SEQ ID NO: 144:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1944 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

AAAAAGCAA CTATAGATA ACGAGAGCC CTTTGTAAA ATTAAATGCG ATTGAGTGT 60

TTTACGCTCT AGCTGTTTAA CTATAGATGT AACATATGCT GCGTACCCAC CTCAAAATGT 120  
CTGTATGCTA AAGAGGAGCTT GAGCTCTGCG TTTCATATTT CAGCTTTGCG CAGATGTGTA 180  
GTCCCAAGA AAGCATGAG TGCAGATGCG TTAGGAAATG AACGTGCGTC TGCATATGGG 240  
ATGAGACAA AGGGGTACA GCAATGCTGC TGCCTTACCG TGCAGTACG GAGACATCC 300  
AAGCATGAT GTTCTTGCCA CCGGGTGTG TTGTCTGCG TGGCTGTCA TGTCTGTGCG 360  
TCACTTTAT TCTTAAACT GAGGTTTAC TGCATCTGCG TACTGAGGCT AGAGCCACA 420  
CGAATGAGG GTTGGCGCTG TGGCCCCGA ACTAGGGGCT GTGGGTTCAT CACAGTGTG 480  
CCTTTTGTCT CATTAAATA GGTATCTACT TTATAGGGA ATTGTCTCTC CCAATTAAT 540  
TTGCTTAC TGGTCTTTT CTTTTGTGCG AGTATTCAG TGGTATAGCT CTGACAGAG 600  
TCAATTTG CCAAACTGA CAGTGTGTG CTGATCTTC CTTTGCGAA CATCGAGGTC 660  
AAGATTCAG ATAGCCCTTC CTAGGCACT GACCTTTCTG GCAATGGGAG TGTGTTTCA 720  
CAAGTATTT TCAATTTAC TGCAGATGT CCAAGGCTG CTGTAGAGCT GAGGTGTGTT 780  
CCCCGTGCG TGTGTACAG TGTCAAGG ATACATCTT AGGGTGTGAG AATCCAGAG 840  
CAGAGACTC CTGTGTTGT GTTCGAGAT CCTGTACTA GAGGGTCTG GCGAGAGAA 900  
CAGACAGCT TTTCAGAT GAGCGGAG GAAACAGTG GTTGTGCTG TGTCTTACT 960  
GTCTGAGAC TGTGCTGTG GGCATTTGAA AAGTTAGATC TGTGATCTCT GGGGTTTTG 1020  
TGCCTTGT TCAATCTTCC ACTGTAGGCG AAGCAGACA GTCTATCTC TCCAGAGCT 1080  
GCTGACTC CAGGTAGAC TGCATAGAG ATCTGTAAAT ATTGTATAG AATTTCTTG 1140  
GTATGATAC ATTGTAGCTG CAGGTGACA ATTGCCAGG TACATGTAA GCTATTAAC 1200  
AGTCAATCTT AAGAGGAG GATGCTGTC ACTCATGGA TCAATGAGTC CATGCTGCT 1260  
TGCAGTTC CTTTTCTT CTTCAGGTT TGTCTTTC TGTGTGTGCC CCGACAGGG 1320  
AAGACTGCG GGTGTAGTG GAGAGACATA TTAGAGATAT AGCAATGGA CCGAATGCG 1380  
AACATGTGAG AGCTTACTGT GATATGAGAG AGTACTGCT GCAAGCTGCG GAGGTACGT 1440  
GTGATGCTG TATGTGACA GGAATTAAT ATCTGAGCT CTGAGCTT CAGCTTCTCG 1500  
TCAATGCTT CCTGTATAC TGCATGGA CAGATTTGA GTTGCAGTC TGCACCTTAA 1560  
ATCACAGCT GGGCATGTC TGGCTGTAG GCTGCGAG TTCTATCT GTTCTTACT 1620  
GTGAGCTGAG ATGCTGAGC TGAAGGTTT TCTGTGCTA GACTGTAGC TGTAGCTT 1680  
GCGCTGTGTA GAGCCCCCTC CTGTGCGCTC AGTGTGCTC GTTGTATAC ATCATCAGA 1740  
AAGTGGAAA GGTGAGGAG AATTTTCTG CCGTACAGAG GTTGTAGAG AAGAGACA 1800  
GTATTTGAT GAATTAACA TATATCTTG TTTTTCTCA ACGAATAAT TATATGAGC 1860



AAATCTACT GCTCAAGTT GATGTGTTA TTCAATATA ACTGTAAGTT TCTGATTATA 1920  
AAAAAAAAA AAAAAAAAAA AAG 1944

5

## (2) INFORMATION FOR SEQ ID NO: 145:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1021 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

TCGACCCACG COTCCGGGT GCGACGGG GAGTTCGGG TCGAGACCG TGCTCTGGG 60  
CGCGCCCTTC ACCATGGCT CGCGAGCT GACTACAC ATCGAGTCC CGGATCAGC 120  
CTCTGGAGC CAGAGAAC GCGCCAGCC AGGTGGAG GAGCGAATA CTGGGAGC 180  
TGTGTGATT CTYTGGCT GGGTGGCT CAGGACAG AACTTTCCA AGTACAGTC 240  
CATCTACAC AAGAGGGCT GCATGTAT CCGATACCA GCGCGTGG ACATGCTTT 300  
CTTCTCCAG TCACTGGTA TCCCTTCACT TGCTGTTTG GCCAGAGC TGCTGAGCT 360  
GCTCTTTGAT TATGAGTNG AGAAGAGGC CTGCTCTTC CATGTCTTCA GCAAGGTGG 420  
GCTCATGCTG TACCGTAG TCTGGAGCT CTGCGAGCC GTTGCTTCT GCGCGCTGG 480  
TGTGTGGGC ACCATCTTG ACAGCGCTC TGGTAGAGC AACTGGTAG GGGCTCTGG 540  
GGCCCTGGCA GGCATCTGG AGCGCGGGC CGCATCTCTG GCGCTGTGG TGCTGTGGC 600  
CTTTCCTGCA GTGTGCTCC TGTTCAGT CTGCTTCTT CCAATCAGC CACTCTTCA 660  
CACCCACTTC TATGACAGC TACAGGAGC GGGCTCTGC TGGCCGAGC TGTACTCTA 720  
CTCAGGGCT GAGGAAGTAG TCTTGGCAG AGCATAGAA CGCATGGTGG AGGCAAGCTT 780  
GGCAGCGGG GTCTGGGGC GTTCTGTGA TTTGTGTCA TCTGACAGC TACGCCACT 840  
CGTCACTAC CTAATTAAT ACAGAGCT CTGTGTGAC TTCATGCCA ACTGGTGG 900  
CTCTGAGC CATCTGCA TCTCACTCT GCTCGAATA TAAATGGCTG ACACCTGCC 960  
50 AAAAAAAAAA AAAAAAAAAA ACTCGAGGG GGGCCGGTA CCAATTCGC CATTANAGG 1020  
T 1021

55

## (2) INFORMATION FOR SEQ ID NO: 146:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1285 base pairs

60

- (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

GGCAGGGA GGGCAGGC AGCATCGG CTTTCAGTT CGGTCTCTG GTGTACGGC 60  
AAGCCAAAT AGGGAATGC GTTCCCTCA GTCCAGAC CTATCAGAT TGGATATGC 120  
CTTCATATT GATTGANTT ACAGTGGTT CAGAGTGTG CTACAGTTT TAGGATTATA 180  
TAGAAGACT GGTAAACTG TATTCTTGG ATTGATTAAT GAGAGAAAA CAACTTCT 240  
ACATGCTA AAGATGACA GACTGGACA ACATGTCCA ACATTACAT CCACTTCCA 300  
AGACTGAC ATTCTGCA TGAGTTTAC AACTTTGAT CTGGTGGAC ATGTTCAAGC 360  
TCGAGAGTG TGGAAACT ACTTCTGCT TATCAATGGC ATTGTATTTC TGTGTGATG 420  
TCAGAGCAC GAAAGCTGT TAGAGTCAA AGAAGAACTT GATTCACTTA TGACAGTCA 480  
AAGCATGCT ATGTGCTTA TACTGATCT TGGGATTAAG ATCGACAGC CTGAGCCAT 540  
CAGTGAGAG AGGTGGAG AGATGTTTGG TTTATATGCT CAGACACAG GAAAGGGAG 600  
TATATCTCTG AAGAACTGA ATGCCGACC CTTAGAGTT TTCATGTGTA GTGTGCTCA 660  
AAGACAGGT TACGAGAG GCTTCCGCTG GATGGACAG TACATTGAT AACACAACT 720  
CAATGTGTT CGAGTCTCA AGTTCAGGC TTACTCAGAG ATTGTATTC TCAACATCA 780  
TAACTGAT TCATAGACT TTCTGCTGT ATAAACAGA TGTTTTTTAG ATTATTATA 840  
TTAATCAAC TTAAATTGA TGAGATTGA AACTGATTC AAGTAAGTTT GAGTATCA 900  
ATGTAGCTT TCTAATCCA TAAAGTACT TGTTTTTAC AGTTTATTA CTGACATCA 960  
CCAGGCCA TTGTANAGA GCACCTTTC ACCAGTACT TTGAGCACT TTTTACAGC 1020  
ATGAAGCTAT AAACATAT TAAAGCTCA TCAATTTAA TTTTTTAT ACTTTCTCG 1080  
AACTAGTTT TAAATTTAG ATTATATGC CACTATCTT AAGTGTACAG TTAATATA 1140  
GCTATTCAA TCAATGATG ATGCTTACA GTTTGATA ACTTTTTTC TTATCGAAC 1200  
GTATGCAAT AAGACAACT CTAATGTTG GCAGAAAAA AAAAAAAAAA NTCAGGGG 1260  
GGCCGTACC CAATTCGCC TAAG 1285

## (2) INFORMATION FOR SEQ ID NO: 147:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1386 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60

## (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5	GGGACGAGGT GGGCGAGGGG TGGTGGTTC TCTGGGCTC GGGACGAGGT GAGGACCTTC	60
	ATGAGAGCCA GGGTCTGGAT GGGGACCTTC GGGGGGTGCA GAGAGTGTG GGGGGGCTTC	120
	GGAGGGGGG GGGAGAGCGG TTTCAGATTC TTACATGAGC TTGAGCAGGT	180
10	TTTCATATGA TGGAGAGCCA TGGCCTCTTC CTTCAGATAT TCTGGATAT AGCAGATCA	240
	TGAGTAGAGA GTTGGAGAAA GAGTGCAGAG GCGTCTTTCG CAGCTATTC GCAATTGAGA	300
	TGAGCCGACA GGGAGCGTTC AGAGAGAGC TACCTCATAT GGTGAGATG TGGAGCAGAG	360
15	GCGAGATCT CATTATGAG CAGAGATTC AGAGTTTCA GATTAGCCAG GTGGTTAGAG	420
	AGTCCATATC AATCTCTAGG GAGGGATATG AGACCTTCTT CAGACAGCTC TACCATATCA	480
20	AGATCCCCC TTTCATCTTT TCTGGGGCCA TTGGTATAT CTTGGAGAGA ATTAATCCAG	540
	AGATTGAAAT GTTCCAGCCC AGCATTCACA TCGTGTCTAA CTACATGAT TTATATAGAG	600
25	ATGGTTTCT CAGAGGATTT AAGGGCCAGC TGTATAGAGC ATAGAGAG AGACCTCTTC	660
	TGTGTAGAAA CTGTGGTTTC TTTCAGAGAC TTGAGGGGAA AACAGATATC ATTCCTCTGG	720
	GAGAGCTAT GGGGAGCTTC ACCATGGGCG ATGGGGTTTC TGGTGTGAG AACATTTCA	780
30	AATTTGGCTT CCTGATATGAC AAGGTGAGAG AGCGGGGAGA AGCTACATG GACTCTTAGG	840
	AGATCTGCT GAGAGAGAGC GAGACTCTG ATGTGTGAA GGGCTATCTC CAGGAGATTC	900
35	TGTCCAGGG GGTCCAGCTG GAGATGAGAG GCGCTGAGAG GCGCAGGCTG CCAAGCCGTC	960
	TCCAGGCGCT GGTGAGAGAG GCGGCTCTCC CAGATTTGTC TCCCTCTGTA AGCAGAGCA	1020
	GAGGCCAGG TGGCCAGAGG TGGCTGGGTC CTTCGGGCCC CATTCTGCT CATTTCCTTC	1080
40	AGCAGCTCA TCAACAGAGG CTTCAGAGAA CCGGCGCATG TGGCAGGGCA CAGGCACTAT	1140
	TCTGTATGAA CATTGAGACA CAGCATGCA GTGCTTAGG GATTGTCTAC TCCAGGATTT	1200
45	TTCTTCAAAA TTTTATATCA TGGAGATTC AAACATATAT AATGTGTAA AACATGAAA	1260
	ATTTTAAAA TGAAGAAAA GCTGCTTGA TTCCAGGGAT GTGGTGTGG GTAGAGCTTC	1320
	GAGCTTTGG CTTGGGGGCA GATGGGATTC TTTCAGAGC AGATTTTGA AGCAGCAAA	1380
50	AAAAAA	1386

## (2) INFORMATION FOR SEQ ID NO: 148:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2098 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

## (D) TOPOLOGY: linear

## (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

5	AGGCTCTTTC CCGGCGCTTG GAGCTGTAC ATCTTAAAGC TCCAGGGTTC TGTCTTTTTC	60
	TGGGTAGAGC CATTGTCTTG ATCCAGGTTT CTTGAGACTG AACAGAGAGG GGGCTGTGAA	120
	CCCTTAATGG GTTATATATC CTCCAGGCA GGGCTGTAC CTTGAGATG GAGTTTACTC	180
10	CTGGCAGTGG TTGCTGAGCA TGGGAGAGC AGTGGAGGC ACCGACTGT GTTATCTGGG	240
	CCTTCATATG AGTGAAGACC TTGGGGAGAA GGGGCTGTGG ATGAGAGAT GAGCTGTAGC	300
15	CTTGGGCTAG CAGCTGGGCT GGAATCTCT GGTTCATATG ACTGTATATC CAGAGAGAGA	360
	AACTGTATAT CTGAGGATTC AGATGTAC CAGCAGCAAA GATTATTTT CAGAGAGACA	420
20	GAGCTAGAAA TGGGCTGTTC TGGCATTTCA GAGTCAAGCA AAGCAGGAG GGCAGAGAGG	480
	CTTCTGTGGG TCTACAGAGG AAGGTCTCTG TGAAGGCTAT GAGTTTATTC CTTCAGCTTT	540
	GCTGTAACT TTGGGGCTTC GGGAGAGC TGGCAGATC CCGTGGGCTT GATGGCAGTTC	600
25	TGTGCAATCC TGGCTTTTTC CCGAGCTTTC CAGAGTCCC TCCGAGCAG ACCTTTCTTC	660
	GAGCTTCCCT GTCCAGCTGG CTGATTTAT GTTCAGAGCA CTGAGACTGC AAGGGAGAGA	720
	CTTCTTCCAA CATTAGAGTG AGATCTTGC CCGCTATCT TGGCATTAAA AACATTTTAA	780
30	AGCATAGAGC ATGTGCTTCT TTGGCCTTCT CTGTCTGTT CCAATCTTCT GCTTCCAGAT	840
	CAGTCTCTGG GAGCATATGG ATTCAGTCCC CCGCAGCTCT GTGGGAGAGC CATTGTCTCT	900
35	GTCAATCCAG AACCTGCTCT GAGGTGAGAG GTGAGAGAG ATTCAGAGCA GCGTGAATCC	960
	TCCAGGCCCA CTTTGCTTAC CTTTCCCTTC CCGTCTCTCT TCTTTCACA CAGCAGACTT	1020
40	AGCTTTTTC ATTCAGATTC AACATAGCCC CATTCTTTAT GAGCTGAGAC CTTCATTTTA	1080
	CCTTGGCTAG ACTGTAGAGC TCTGAGAGGA GTGGGGTCTCA CTTTCTTCTT GCGCTATGCA	1140
	GTGCAAGCTT CAGTCTTAC CAGCAGAGGT TGAATCATCT GCTTCAGAT CTCTGGGCTT	1200
45	TTGCTTTGAC CATTGAAGCT AGCTGGGCTG GTCTTGTGTC CAGCTTCTT CCGCTCTCTC	1260
	GGAATGCTCT TTGGAGGGCC CTGTCTTTCC TCCGAGACA GTTGTCTTGG CTGCGATGGC	1320
	AAAGTCATCA GGGGCTTTTA CCGTGTTCAG CAGAGATAT AGCAGCTGCC TGCATTTAT	1380
50	CTTCAATGG TCACTGAGAG TCCAGAGCTT TCAATCCGCA TGTTTTCTTC GGCAGAGAGG	1440
	GCTGAGATTC CATTCTGTGG GAAATGGGCT GAGCTGTGAT AACAGACTAC ACCCGAGTGG	1500
55	CTCTTTCAT GATGTGTGCT GCTCATGTTG CTGATGAAA CTAGAGAGCT AAGCTGCATC	1560
	TGGAGATGAC TTTCGCTGCA GAGGTGCTTT GCTGTCTTTC AGATGATGTC ACTGTGTTCC	1620
60	CTTCCGCGCT CTCTTATATC CATTGGTATT TGGAGCTGTC CCAAGTACTT TGGGGTCTCA	1680

GCTGGAAATTC CTTTGTGGTT TGTCTCTCTG CTCTCACTC TTGTATTAG AGGATTCCA 1740  
 CAAGGGGAGA GTGGATGCC TGTCTCTCTT GTGCCAGACC AGATTCTTCT GAGGGGGCTT 1800  
 5 GACCTTAACC CTCAGCTCA GCGCTGTACA CCGTGAACCTG TAAATGAGTG GGGTTTCTG 1860  
 ACTGTAAATCC CTGACACAG TAAACCAAA AGGACTCTTG GGGCTCACT GTGAGAGCCA 1920  
 GGGTACTCTA CTCTGCCAG TGAGACAAA CTGCTAGGCT GTATCCGCTA ATTTCAGAT 1980  
 10 GAGAACNTT AACNTAATA ATTGTAGTA AACATAACT CATGAGACT AAAAAAANA 2040  
 AAAAACTVGG GGGGGGGCCC GTAAACCATT GGGCCCTTNG GGGGGGNTT TTAAATTT 2098

15

## (2) INFORMATION FOR SEQ ID NO: 149:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1847 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TCGACCCACG CTTCCGAATC GAGCGCGCG CGCGAGCGCG TTGKGTCTG GTCTTCGGT 60  
 30 CGGCCCCGCG GACGAGAGCG TGCCCCCGCG GCGGGGAGAA GATGTGTGCK AGCGGCGCTG 120  
 GCGCGCGCAC GCGCGCGCAC GATGTAGGCC AGCGCGAGCG CGGGGTGCG CGGACGAGCT 180  
 240 GCGCGCGCTG GCGCGCGCGCG GCGGANTGCC GCGCGGGCGG GGGTGGAGCT GATCAGAATA 240  
 300 ATGTTTACGA TCACCCCTCT GCGAGAGCTG AAGGTGTACA TCAGGAGTGC GCGTCCCTG 300  
 360 GTGGCTTTCA TGATCAGGCT AAGGCCGATG GCGATAGCTT TCTGACCTT GGGCTACTTC 360  
 420 TTCAAATACA AGGAGATTAA ATCCCGCAGAA ATGCGCAGAG ATTGGAAATAC TTCTCTCTA 420  
 480 CGTTTCAATG ATTGTGACTT GTGTGTATCA GAGAAATGAAA CCGTCAAGCA TCTCACAAAC 480  
 540 GACACACAAA CTCGGGAAG TACATATACC AGCGGGGAGG CCGGAGCTTC CACCCGCTCC 540  
 600 CCGCAGGCCC TGGAGGACTC GGGCGCGGTG AATATCTCAG TCTCAATCAC CTTAACCTTG 600  
 660 GACCCACTGA AACCTTCTCG AGGGTATTCC GCGAGCTCA CCGATCTGTA CTCACACATC 660  
 720 TTAGGGGATC AGATTGAGCT TTCAAGGAGG GAGGCCGAGG AGGAGATPAA CATCACTTTC 720  
 780 ACCCTGCTTA CAGGTGGAG CTCAGATGAC TGGCGCCCTCC AGGTTACTG TGAGCAGTGG 780  
 840 GTATTTCAG CCGTCATGAC CCGTCAAGCC AGCCCTTGCGG TGTTCGCGGT CACTGTACAG 840  
 900 CCACCGCACT GTGTCTCTGA CAGTACAGC AAGGCCAGCG TCTGGTACAA GATCTTCAAA 900  
 960 ACTGCCAGAG ATGCCAAGAC AAAAAAGCCC CAGGATTACA ATCCTTCTG CTGTTATAG 960  
 1020 GGGGCGATTC GAAAAGTCTA TCATGCTTTA AATCCCAAGC TTACAGTGTAT TGTTCAGAT 1020

60

GATGACCGTT CATTAATAA TTGGCATCTC ATGCACACCA GTTACTTCTT CTTTGTGATG 1080  
 5 GTGATACAAA TGTTTTGTCT TGTCTTTATC AAGGGGAGAC CTAGCAAAAT GGGTCAGAGC 1140  
 AATCTCGAAT TTGTGCCGA GAGGTGGCT TTGGCTGAGG CTTAATTOCA CAGCTCCTTG 1200  
 TTTTTCGAGA GAGACTGAGA GAACCAATAT CTTTGGCTGC TGAACCCAGC CTGGGCGCTGG 1260  
 10 ATGCTCTGTG AATACATAT CTTTGGATGT TGGGTATATC CAGCCAAAGA CATTTCAAGT 1320  
 GCGTGAATCT GATTGTGACA TTTTATATA AATCTATTTCA GAATTTGCTC CATTAATGCA 1380  
 CCGCTTTTCC CTTGGGTACA GGCAGAGCCC TTCAACCCCA CTTTGGACTT GAGGACCTAC 1440  
 15 CTGATGGGAC GTTTCAGCT GTCTCTAGAG AAGGATTCCT GATCTAGCT GGTCAAGAGG 1500  
 ATGTTTTCAC CAGGCTCACA GAGGCTTGG GTGGCTGATG GGGTTCGAGT TTGGTTTGGT 1560  
 20 TCTTGTATCA GCGCAATATG TAGAGACAT TTGAACACAGT CTGCACCTTT GATACGGTAT 1620  
 TGCATTTCCA AAGCCACCAA TCCATTTTGT GATTTTATG TGTCTGTGGC TTAATATACA 1680  
 25 TAGTACAAAC AATATATCTT TTTTCTCCAT TTTGCTTCCA GGAACATAC CTTAAGTTTT 1740  
 TTTTGTTTTG TTTTGTTTT TTTTGTTTT GTTTTCTCTT ATGAGAGAAA AATAAATAG 1800  
 TCACATTTTA ATACTACCAA AAAAAAGTGC AGGGGGG 1847

## (2) INFORMATION FOR SEQ ID NO: 150:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1569 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

GACCTTGACG AGAGAGAGCC TCTCTCTTCA GGGTGTGTGC TGTGTTCAG TGACCGTGGC 60  
 120 GGATTAGCCC AACTGGATC CGCGGGTGT GAGGTCTGGA CAGTCAAGA AAGCGGTAGC 120  
 180 CAGCGCTGTT CAGCAGGAG TAAATCTCT TGTGGCTTG GAGCGCTTC AGGTTCCTGC 180  
 240 AGAGGAAGCT CTTTCTGGG CTGGTGAGCC CTGTGACATC ATGACAGCA GTGATGAGAT 240  
 300 GGATGCGCAG GAGGAAGCA TCCATGAGAG AACTGTCTCC AAAAAAAGA AAGCAAGAG 300  
 360 ACACAAAGAA GAATCGAGG GGGCTGGAGG AGAAGATAT CCAATGGATA TTGGCTATTT 360  
 420 GCTGGCTCC TATATCCGTC CTGAGGACAT TGTGAATTTT TCCCTGATTT GTAGAGATCC 420  
 480 CTGCACTGTC ACTTGCAGT CTGCGTTTGG GACGAGGTG TACCGAAGCA CTACAGCTGG 480  
 540 GATGCTTCCC TCCCTTTGGG TCTGGGACCA GATGTCAATG AGAAGCTGG CTGTCTCGGG 540



5 ATCTGATCA ACAGGCTGTC GTTTCAGGAA TGATTAACAGG AATTGAGGAA TTATCAATG 240  
 GTCTGGAGCC GGGCTCTAT GGAATCATATT TCTATATATT CCAATGTGAA CTTAAAGAAC 300  
 TGCCAAATAC AGGACAGAC TTGGGAACAA ACACAGCCCT TCAGACAC TCATTGACAGAA 360  
 AATCCATCAT CCTGGCCCT CCTCTCTAT TTGGAGCCCT TTCACTACTG CTGGCTCTGC 420  
 TTCTGGCCTT GTTTATTCGG GAACATACCA ATTATAGCTT AAGGTCCAG AGTTGAGAA 480  
 10 AGCACTGTGG CAGTCACAGC CACTCTCAT ATACACAGCC GCGAGAGAG GCGAAGAGAC 540  
 CTTTACTCCA GGAACAAAT GTGTGAGAC TGAATTCAG AGATTTTTC TATCAGCACC 600  
 15 CAGGCTCTTAG TTTCACCTC TASTTCTGGA TGTATATCC ATTTCATCC ACAGTGTACT 660  
 TTAAGATTTG CTTAAGAAAT GTATCTGCAT GAATCTCCTG GGAATTAAG GAATGAGAA 720  
 20 CTTAGACCA GACAGTTTC CAAAGATGTT ACAATTTCTT TTGAAAACCT TTTCGTTTAT 780  
 TAGACCAAT TTCTGCCCAC TAGCTATTT GTTTTATAT ACATCTCTTA ATTAAAACT 840  
 ATATATGAA CTTCTTAGAT ATTAGCAAT GTCTCTGCTA CCATTTCTCT AAGGTGTTCA 900  
 25 GCTTTACTC TATGCTGACT CAGTGAGCA CAGTAGGTAG TATGTTGTG GACCTATTTG 960  
 TTTTACATTT GTAAATTTT GAGTCAGATT TTATATTTGT AAATCTCTGG GTCAAAATAT 1020  
 30 TCAAGCCTT AATGACAGTG CACTAAMCA AAGAATGTT AATGAAATG TTTCATTTA 1080  
 AAAAAAAAA CTTTAAAGAA AACTGTACTA AATCTGAATC ATGTTTGTAG CTGTGTTGCA 1140  
 GTACTTTTAA ACATTTTCA CTACTTTTT TCAAGTGAGA AAGTATCAGC CATTATCAT 1200  
 35 TTAAGTTGG GTATTGAGG CCGTAACT AATGCTGGC TCAAAATAT TCCCACTTA 1260  
 CTTCTATAC CACTATCTT TTATGTTTG CATATATATA AGCACTCAA CACTTGAAAT 1320  
 40 CATATCTTAA AATATATTA GTAAAGCTGG TAGCCTTGAA AATGCTAGTG TGATATCTAT 1380  
 TATGTAGATA AATATATATA GTGGCTTTC AGCACTGTCA CAGTAACACT TTATTTACAG 1440  
 45 AGCTAATGTT TGTCTTAAT TTTCAGACC CTAGAGGAGA GCTTTATACA ATTACCATG 1500  
 TGATTTCTC TAAAGTGTAT ATTTTGTGT CCAATTATAT TATTTAAAA AGTGTACTT 1560  
 50 TGTAAATTT GTATATNAG AACTGTATAG TTACACTGT TTTCATCTTG TGTGTGTTA 1620  
 TTGCTTAATG CTTTITTAAC TTGGAACACT CACTATGTT AATTAAGTTC TTAAAGAA 1680  
 55 TGTAAATATT YGTATATTA AGTTAATAT TTTAATAT 1719

(2) INFORMATION FOR SEQ ID NO: 153:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 863 base pairs  
 (B) TYPE: nucleic acid

5 GGACAGAGG AAGCCGGAC GATGTCCCA TCACAAACCA CATTGGAGTT TGGAGTGCT 60  
 TCCCTTAGAG CAAGGAAAC AGCTCTCAT CAAGGAAT AGAAGCTCT CCTTAGTGG 120  
 10 TAGGAGACA GCGAGAGCG GTTTCTGGG AACTGTGGA TGTCCCTTG GGGGCGGAG 180  
 AATACAGAG GAAGATGCT CAGACAGTA ACTACAGCT GTGCTCTCT CTGACGTTCC 240  
 TCTGCTGTC CTATGACCT TTGTCAATT CCTTCTAGA ACTCTCCAA AAGACTCTG 300  
 15 TCATCAGCT TGTCTCTTC ATCATCAGG ATATGCAAT CCTCTTCAAC ATCATCACTA 360  
 TTTTCTCAT GTTCTTCAC ACTTGTCTT TCCAGGCTGG CCTGCTCAC CTCTATTC 420  
 20 ATAGTTCAA AGGACCAATC ATCTGACAG CTGTGTACTT TGCCCTCAGC ATCTCCCTTC 480  
 ATGCTCTGCT CATGAACCTA CCTGCAAAA ACTCCAAAG CTTCTATGCG ACAGATGAC 540  
 TTCAAATGCT GTTTGTATTC CAGAGCTAG CAGCATGCTT GTACTCTAC TTCTATTAAC 600  
 25 GCAAGCCTT AAGACTAGC GATCTCACT TCTACAGGA CTCTTTGTGG CTGCGCAAG 660  
 AGTTCAATCA AGTTCGAAGG TGACTCTTG TCACACTCAT GATACTTTT CCTTCTGGA 720  
 30 TAGAGGCCA CATTTGCTGC TTTCGAGGG AGATTTGGGC CCAATCCATG GCGCAAAACA 780  
 GTTGGATTT TCAAGGAAA GCGTTCAAG TTAGGCTGT TGTTCAGCC ATTTCCNAG 840  
 35 AAGGGAAGG GTTTCCCTTTC CCT

(2) INFORMATION FOR SEQ ID NO: 154:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1101 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

50 AACACAAA AAGATGATTT TCTTCTGAAA TTGTGGAACA TGAGGATTTCA AGTTTTTATT 60  
 TTGTACTTAG GTCTGGAGG AACATCCGAG TTCAACAAAG CCCCCTCTCT TCCTCTGGAG 120  
 CCGAGCCTTG CCGTGGATC AGTCCCACT GAAACATCG AACAAATAG AGAAGATTA 180  
 55 GAATAGATG AATGACCCA AATATAGGTT TTCTTGGGCG AGCATGTCTT GCAATAGCAA 240  
 AGGTGACATG ACACAGGAG AGCAGATGCG CACCCACAC AGAATACAGT GTGTGTTATT 300  
 60 ACAGAGGCC AGCATTTGAG CTTAGAGTCC TTCTACTTAC CTGCTATTGG CATTGAGGT 360

CGGAAAGCTT CTAAGTCCCG AAAAGCGGAA AGAGAGCA GTTCCCTTAA 420  
GAAGTCGAG CAGAGCTTGA GAGCTTATGT ATTGAGCTGA GTGAGAGAG TACATGATGC 480  
TGTCTGCTTT GAAAGAGACT TTTCTCTGCT AGCTGAGCTGA CTCCTTCCCTT AGTTCAGAGA 540  
AGAGTCAGAG CAGAGCTCTG CTGAGTACCT CTGTGATGAGC AAGAGCTTGG TTATACAGAG 600  
GTGATTCGCA GGTTCGAGAG TTTATGTGTC CCGAAGGCAA AGCAAAATAT TACATTAATA 660  
ATCCAACTTT AATGATATAC ATTAAAGAGA AAAAAAGCAA AAGTGTAGAGC 720  
CAGAGCTGCT GAGTATTCGA GCTGCAATTTG CCGAAGGAGA ATCCAGAGCA AGTCCCTCCG 780  
TGTATTTTGT TCTTGAGAGAG GGTTCAGCTCA GAGCTAGAT CTATCCAGCA TGAAGAGAGG 840  
CAAGCCAGAG GTTGTCTGCA TCAAGAGCAA CCAATTTTAA GAAAAAGAGA AGAGTTCCTT 900  
AGATGAGTAA TTTGTAATGA AGATGATCAG TGAATACAGC TGAACAGATG CTATGATATC 960  
ACTATGTGTC CTTTATAGAA TAAAGATATC ATATCATCAT TCCCTTGGGG AAAATTTGTA 1020  
TTCAAGTATA AAAACAGAG ATTATTAATA AAAATTAATA GAAAGCTTAA AAAAAAGAG 1080  
CTCTGAGCGA ATTCCTCGCA G 1101

30 (2) INFORMATION FOR SEQ ID NO: 155:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2011 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1.1) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

40 CAATTACCC GTTGAAGGC TAGGTGTATT GCGAAGCCCG NAGGCTAAGG TTTTAATTCG 60  
GAGAGGCGAA GAGCTTGAAG GAGAGGAGAA GCGAGAGGTA GCGAGAGGCT AGCAGGTGAG 120  
TTCCAGAGGC TGAAGAGTTT AGCAGAGGC TGTGTCAATG CCGTGTGATC AAGAGAAAC 180  
CAGAGTCTC CTGAGTACCT ACCAAGCTTG GTTGTACAAA AAGCAAGGTG GAGAGCTATT 240  
TTGTGACATG AGATGATCA CACTTACCTG TGGGCGAGTA TGTGAGATG AGTGTAGAGTT 300  
GTTTACATG AGTCCCTCC TGGCCAGCAC AATTTGTATA CATAGCTTTC AGAATGATAC 360  
CAAGCCCTTC CCGAGCTCC AAGCAAGGC TGGTCTGAGC CTGTGTATAT ATGTCAATTT 420  
TACGTTTTT AATATATAGC TTTGATTTCT GTTGTGTGTA TTTTGAAGCA GTGTATGAC 480  
CTTCAATTTA ATATCATGT GTGATACAG ATACCATAT ATGTGTGTC GTATGCAAT 540  
ATCTCTATTC TGTAGTTTCC AAGAGTTCAG CTGAAGGAGA TGAAGTCTTC GAGCCAGAGA 600  
GAGAGCTGAC ATCCCTGCTA ATATGTGTTG CAGAGATAT TACATGATCT TCCATTAATA 660

TATTTCTATT TCGAGAGACT GAGAGAGGC TGAATGATTT TGTGTGTTCT TTGCGAGCTA 720  
AGTGAAGGTC TGGAGATGAG TTGCTGTGTT CTTCAGAGTC CACTTGGAG CCAATCTGTC 780  
AGATTAAGC CCGCTTTTGG CTGTGTGATA CTCCTGCTCT GCGTGTGAT GTGTGTGATA 840  
GTCACTCTTG CATGCTTCC ATGTCTGATT TGTGCAATTT GGGGATTAAT GCTGAGAGAG 900  
AGCAATTCGA GTTGTGTTGA GAGCTGCTTG CCAATGATAG ATCACTCTTG TTGAGCTGAT 960  
ATGTCTGCTT GCTTCTGCTT TTTCTGCTGA AAGAGAGAGC AGCTGTGCTCA 1020  
GGCCAGAGCT GAGTGAAGAG AGCTGAGACT GGTTCATGCG GTTCTTGAAG CAGAGAGAGG 1080  
AGATGTGAT TTTACTAAT TCCATTAACA AATTTATGTC AAGCAAGACT CCTTGCAAT 1140  
CGAAGAGTC AAGCAAGATA GAGTTATGAC TTGCTCAGAG ATATATGCT CTATAGGGGCT 1200  
AATTCAGGC ATCAAGATT CTGTTCAGAG GCGAGAGTA GCGTGTATAT TGTACAGACT 1260  
TTTATTTTAA GCAATTAATT TTGACTTGT TCCCTGAGAA GTGCAGAGAG CATTAGACTT 1320  
TCCCAATGT AACTAGAT CTGAGAGAG CAGAGAGCTG TATATGTTCT GTTCCGAGAA 1380  
GAGAGAGAC TTTTGAAGAC CAAATGATCT TAAATGTTAT TGGCCAGCC TGGCTTTTCC 1440  
GAGTGAAGAA TTGCAGTGA GATGATGTT TGAAGAGAG TGTGTGAGC CATGAGCTTAA 1500  
CAGAGAGAGC TACATTAATT CAGATATGTC CAGAGAGC AAGCAAGAG CATTACTTGA 1560  
CAGAGAGAA ATGATGTAT TGAATGTCCT GTGTGTGCAA AATGAGAGCA CCAATTTCTT 1620  
TGAATATGTC CCACTGAGAG GCTGAGAGAG GTGTGTGACA CCGTGTATATC CCGAGACTTT 1680  
GAGAGAGCTGA GAGGAGAGAG TCAAGAGAT GCGAGAGTTT GAGAGAGGC TGAAGAGAA 1740  
TGGAGAGAAC CCAATCTTGA CTAATTAATA CAAATTTAGC GAGGCGTGTGT GGCATGCGGCC 1800  
TATATCTGA GCTATGAG AGGATGAGAG CAGAGATAT GCTTGAAGCC TGGAGAGGAG 1860  
AGATTTAGAG TGAATGAGAG GATGTGTGCA TTGCATCTCC GAGCTTTGGG GCAAGAGAG 1920  
CAAAATTTCC GTTTCATTAAT TGTGTGAGTA TGTATGATAT GCTTATGAT ACCGTGAGCC 1980  
TGAAGAGGAG GCGCGATGAG CAAATGCGCC TATGATGATC GTATTAAGAT C 2040

50 (2) INFORMATION FOR SEQ ID NO: 156:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1981 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1.1) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

60

60  
620  
640  
660  
680  
700  
720  
740  
760  
780  
800  
820  
840  
860  
880  
900  
920  
940  
960  
980  
1000  
1020  
1040  
1060  
1080  
1100  
1120  
1140  
1160  
1180  
1200  
1220  
1240  
1260  
1280  
1300  
1320  
1340  
1360  
1380  
1400  
1420  
1440  
1460  
1480  
1500  
1520  
1540  
1560  
1580  
1600  
1620  
1640  
1660  
1680  
1700  
1720  
1740  
1760  
1780  
1800

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

CTCCACCTT GAGCCCTTCA CCCCCTCCAG TTCCCTCCAG GTTGGCTTC TTGCATTCCT  
TTTCTTGCTA TCACTGTTTG ATTGAGGAA CACCCCTCTC TTGTCTACCC TCAATATGA  
5 GCTCACTGTG GAGGAGCAG TCCGCGACAG CTCGCGCTYA TGTCTATTTT TACCCCTCAA  
GACCCGCAAA ACTCTGTGTG GACACAGAT GACCCCTCAG CCAAGATCCT GCACAGCTT  
10 GTGCTCTCTG CAGCTGAGCC CTGCTCCATG TTAGAGAAC AGCTCATGGA TCCCGCGGA  
CCTGGGGA CAAGGACAT GTTCCGCGCG CCTTGGACA TTACAGACT GTGATTCOC  
420 CTGTATCTTC CCATATCC GCGGACGCG AGCTTGTG ACTGCGAGY TGCCTCTTC  
15 TGCCTGGGCT TCCCTAGCCA GCGGCGGCTC TCACTCTGA TCCCTGTCTT GGTGATGAT  
480 CTTCTCTAGC TCTATCTGAC GCAGCTCAG GAGGCTTTTG GCGATTCGC CCTTTCTTC  
540 TATGACAGC ATGCTGAGCA GTGATTTGT GTCTCTGGA AGCCACAGC GTTCCAGCG  
600 CAGCCCTTCA AGGCTCTCAG CACAAAGGG GCAATGCTGA TGTCTCGAG TGGGAGCTA  
660 GTATGCTGTC CCAATGTTGA AGCACTCTG GAGGACTTTG CTGTCTGCG TGAAGGCTG  
720 GTCCAGACT TCGAGCGCG AGTGAGGCG TCGACTGTGT GATTCAGCT CTGAGCAG  
780 CTGTAGAGG AGAGAGAGC ATTGAGCTC TAGAGCAAGA TGTCTAGAG ATGAGCTTCA  
840 CCTCTCTTGG AGTATGATTC TCACTGAGG GCTCTCTGCG TGAATCTCT GATCATCTC  
900 CAACAAACC CAGCCCTCAAC TTTCTCTCTG ATGCTCTGCG ATTGCGGCG GCGCATGCTG  
960 GCGCATGTAG TCTCTCTGCG CTCACCTCC CAGAGAGGA GTGCGAGCCA GCTCAGAGA  
1020 GGAATCTGAC CAGAGAGATC CATCTACTA TTAGCCCTTG GCTCTGAGCT CCTCGCAT  
1080 TCCCACTCT TTTCTGAGCT TCTTCCAGA ACAGAGAGCG GGTATCTGCT CTGCGAGCG  
1140 CTCTCTCTCC TTCTCTCTGC CAGGACTCTT GCTAGACTT AGCATGCTCT TCACTGCACT  
1200 GTCAAGGCTT TAGATGCGAC CGAGCGGAAA TGTGCGCTTT CTGAGTCACA TCAAGGACAC  
1260 TGAGCATGCG AAGGCGCTA TATCTGTATG ATTAGAGCAC ATTGAGAGAG CAGATTCCTC  
1320 TCTCTGTGTT ATGCACTATC CAGAGGTGGA GACAGTGGAA AAGAACCGAG GACAGGAG  
1380 GATTCGCTAG GTAGAGGGGT CAGGCGAGCT GTACTCACCC AATCTGTAG AGTCTGAAA  
1440 AGCACTGGGG GCTACCTGTT AGCTGCTCTT GCGCTGCGTG TTGCGCGTT CATCTCAGA  
1500 ACTGCACTA CTATGTACTT CGAGTGGGTT TCGAGAGATG GCGGAGACT AGTCTTACT  
1560 CCGCAGAGC TCCAGAGGCG CAGAGAGAG AATGCTGCTT CTTTCTAGT TGTCTTAC  
1620 CCACTTTCTG GTAGGCTCTC TCTCTCTCTT AATCTGCTT GTTTTCTAG ACTGAGCTCA  
1680 AATAGTCCCT CTCTTTAGC CCAATCTCTG CCGCAGCTT GAGTGTACT TTTCTCTCTC  
1740 TGAATATTA GAGGAGTAC TGTCTGTCTA GTTGTGTTG CAGGAGACA CAGTGGCTA  
1800

1860  
1920  
1980  
1981

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

AATCTATTG TTTGAACTC TCAATTAATA TTAATATCA GCTGCGCTG GTGCTCTATG  
CTTGTATTC CAACACTTAG GAGTATAGG GAATCACTTG ASCYAGAGAG TYCTAGACCA  
5 ATCTGGGCAA MAGAGAGCC CCACTCTTTT TAAATTAATA GTTAATATTC TTAATAAAAA  
A  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 157:  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:  
GAATTCGCA CAGCGCGCG CATGCGCTC CTCTTTTGG TCTGCGTGT ACTCTGGCG  
60 GGTCTCTCG CCGCTGTGG GTTGGCCAAG CTCCTGGAG AGATCTCGC TCGAGTTTG  
120 GAGCGGACA ATGCGCTGTT CCGTCACTTT GCTGAGTGT TCCCGCTGA GGTATTTGC  
180 TACCGCCAG ATCCCTGAA CTACCAATA GCTGTGGCT TTCTGCACT GCTGCTGCG  
240 TTCTCTCTG TCATGCGCC ACCGATCTG CAGAGATCA GTAATCTGT CTTCATCTG  
300 CTCATGATG GCGCTATTT CACTTGGCA GCTCTGAAG AGTCACTAG CACTCTGATC  
360 CAGCCATG TCTCTGCG GTTCTCTCTG CTCTGTANT TCGGCGAGT CTTAGCCCG  
420 ACTAGAGG TCGTCAGAC CACTAGAGAG AAGACTCTA GTACATCTA GGAATCTCG  
480 AAGTAGACA TCTCTCTC TTATGCCAT GCAGCTGTCA CAGCAGGAC ATGCTAGAAC  
540 ACAGATCTA TCACTCTGT ACCAGTATA TATCGAGGT CAGCCATGT TGAAGAGAC  
600 ATTCTCTA CCGCGACTG CTCTCTCTT TTAGCTTAC TACTCTTTG TGAGGATAC  
660 ATGTATGCA TATTAACAT CCGTATGTA TATCAATAA CAAATTAAG AGAATAAAA  
720 TTAATCTA CCAAAATCT GATCCGCAA ATAGCACTT TTAATGCTT GGTGTAGTA  
780 TACTCTGAA CTCTTCTG TCGCTTAAA CAGATATA TTTTTTTTA ATGAATATA  
840 AACCATAT CCAATTTAT TTCTCTCTT TAAAGCTTA TAACTATAA MAAAAAAA  
900 AAAAAAAAA CTGCA  
915

(2) INFORMATION FOR SEQ ID NO: 158:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2117 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

5 AGAGGAGAG GAGGTGGAG GGGTCCGAG CATGAACTG GGCAGGAGAG TGGTGGAGCT  
AGTGGCTGAG GGCAGGCTGAG TGGTGGAGAG GGTGGAGAGC ATGAGAGCTG GAGTGGAGCT  
10 GCGCGGGGCTG CTGACCGGGCT AGATGAGCC GGTGTCTGAC TGGCTCTGAG CGAGATGCTG  
CGGCGAGAGG CGAGAGCTTA GCGCGGAGGC AGTGGAGAGAG GATGAGAGAG AGAGAGCTTT  
15 TGGAGAGAGT CTGGAGAGAG AGATGAGCTT AATGAGCGTG TTGAGTTTCA TTAAGAGAGC  
AAAGCCGAG AGAGCTTCA GGTCTGCTT GCAAGGGTGG AGAGAGAGAG GCAAAATTTT  
CTGAGAGAG AGCATGGAG AGAATATTTA CGAGGGTGGT CTAGAGAGTG AGTATGTTCA  
20 CGTGTGTGAG GCGAGATTC AGTTTCCACA TGGTTTCAAG ATGAGAGCTGT AGAGAGATCA  
GTGAGAGTGG TGAATGGAG GAGAGCTGAG TGGCTGTGAG AGGTGAGATG TCAATATTTA  
25 TGAATGAGAT AGATGAGAG GAGAGCTGAT AGATGAGAGC AGAGCTTCC TCGAGTATTA  
TGAAGCTGAG GATGGAGGCT CTGACAGAGA AGCGAGTGTG ATATTTCTCA GCAATGCTGAG  
AGAGAGAGAG ATGAGAGAG TGGCTTTTGA TTGCTGAGAG AGTGGAGAGC AGAGAGAGAG  
30 CATGAGAGCT AGAGAGAGTG AGAGAGAGCT GTGTGTGTGAG GTTTTCAATA AGAGAGAGAG  
TGGCTCTGAG GAGAGAGAGT TAAATGAGAG GAGAGAGT GAAATGAGCT CGGAGAGCTA  
35 CGGCTGTGAG TGAAGAGAGC TAAAGATGAG TATGCGAGTG GAAATGAGCT CGGAGAGCTA  
TGAATTTGAT GAGAGAGTG TGAAGAGAGT GGTGAGAGAG ATGAGAGATTT TCGGAGAGAG  
GAGAGAGT TTTCTGAGTA AGAGCTGAGA AGAGTGTGTC AGCGAGTGG ATTTATTTACTA  
40 CGATGATGAG CATGAGAGAT TGGAGAGAG AGTCACTGCG TGGAGTTGAG AAAAGAGAGAG  
CATGAGAGCTG TTTGAGAGCT CGAGCGGAG GTCTTTTCCG TGGAGAGAG AGTCACTGAG  
45 TGTGCTGATG TGAATGAGAG GAAATTTCTC CTGAGATGTT TTGAGAGAGC TGGTGGCTGAG  
AGAGAGAGAG GAGAGAGAGAG GAGAGAGAGT GAGAGCTGCG GAGAGAGAGC AGAGAGAGAG  
AGAGAGAGCT GAGAGAGAGCT ATGCGAGAGAG TCAATGATTT TTAAGAGATTA TGTTTAAATTT  
50 CGAGTGTGTT GTGTTTCAAG GAGAGAGAG TAAAGTTTAT TGAAGATGAG GTAACTTTAT  
TTAAATGAT TTTTAAAGAT ATGAGAGAGT GGTGAGATTC TAAAGTTTGG GCTGTGTGAG  
TGTGTGTGTT TTTAAAGTCT CATGATGAT AGATGAGAGT TGAATGATCT TTAAGTGAAGA  
55 TGAAGCGAG GAGAGAGAG TGGAGTGTGT TCGAGAGAG GCGAGTCTGAG CGAGATGAGC  
TGAAGAGAG GCGAGAGAGT TCGAGAGAG GTCCGAGAG AGTGTGAGAG GCGAGAGAG  
60 TGAAGAGAG GCGAGAGAGT TCGAGAGAG GTCCGAGAG AGTGTGAGAG GCGAGAGAG 1620

GTGTTAGAT GAGAGAGAG TGAATGAGAG GAGAGCTTCA GCGAGAGAG ATTAAGCTTCA 1680  
AGAGAGCTTC CGAGAGCTTT TCACTTGAAG AGCATGAGAG GGTGTGGAG TGAATTTGAGC 1740  
5 GAGAGTGA AGATTTGTA GAGAGAGCT TTTGAGAGAG AGATTTTCC AGTGTGAGC  
TGGAGAGTCA GTGAGTTGAG CGAGAGTGG CGCTTGAAGT TCGATTTGAG ATTAAGCGAGA 1860  
CGAGAGCTG AGTGAATTC AATTTATGAG AATTTGAGA TATTAATCT TCGTCAATTA 1920  
10 TACTTTATTT TGTAGAGCTG ATGAGAGATC CGGAGAGCT CTGTTAGAGC TTAAGCTGCGC  
CTTGTGAGAG TGGTGAAGAG GGTGTCTGTT TTTTAAATG GAAATGAGTA CGGTTTGAAGA 2040  
15 ATTAATTTT GATGTTGAGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GGGGGGAGCG  
GTAGAGAGAT TGAAGCC 2117

## (2) INFORMATION FOR SEQ ID NO: 159:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2195 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (K1) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

30 TGTGCTTGA TCGCTTTTCT AAAAAAAAAA GAAATTCGAG ATGAGATTTA GCGATGCTG  
TGTGTGAGAG TGTGTTTAT TGCAGAGCTA AATGAGATG ATAGGCTTTT CATTTGCGAG 60  
35 GAAAGCTTT ATTTGAGAG TGAAGAGAG TGTGTTTCC AGAATGTTAG TTAATTTCTC  
GTCTTCTTGT CTGTTCTTTC CTGCTTTTTC CGGAGTGAAG CGGAGAGCTT ATTTGCGAGA 120  
40 CATGAGTGA GAGAGAGCTG TGTGCTTGA CGCTGATTA TGAAGAGTGG CTAGATTTGA 180  
GAGAGAGTGC AGTTTGTGAG TTAATGAGTGT GATTTTGCCT ATGAGATGAT ATTAAGAGAG 240  
45 AAAAAAAAAA GTTGTGCTG TGTGACTGA GTCTTGAAGA AATGTGAGCC ATAGTTTATG  
GAGAGATTTG GAAAGCTTT AGTGAAGAGT GTATTTGAGA ATGGAGAGAGC GAAAGTCCGAG 300  
GAGAGAGAG AGTGTGAGAG GAGTGAAGTGC TGAAGTGAAGA AGGAGAGAG AGTCTGAGC 360  
50 TGAAGAGTGT GAGTGTGAG TGTATGTTGC GTTGTATTC AGAGAGAGAG AGAAGAGAGT  
AGAGAGTTT GAGAGAGT AGGTGAGC AGTGTGAGAG AATATGTTAG AGAAGAGATAT 420  
55 GATGTGAGTA GAGTGTGAG GTTGTGAGTA TGAAGTATG AGAAGAGAG TAAATGAGTGC  
TACTGAGAG AGAGATTTCA TTTGTTGAGC TCGAGATGAT ATTAAGAGAG TGTGTGTCT 480  
60 CGTGTGTTCT AGGTGAGAG AGTGTGAGTA GCTTTTGAAG CGTGTGAGAG GAAATTTGAG 540



5 TTTTCATG GTGACTATCA AAMACGTTT CTGCATGTAC TGAGCCGCA GGAAGAGCT 960  
GGAATCGTTG TCACATATCC TAAGCATCA GTGTTTCTCT TCATTGACAG ACAGCACTTG 1020  
CAGACTCCAA AAMACAAAGC TACATCTTC AGATTATGCA GCATCTGCTT CTACTGCCA 1080  
CAGGAACAGC TCACCCACTG GGGGAGTGGC CACCATAGAG GRTCACTGCC GTCTTATAT 1140  
GCCAGAGTAG AGTACTGACC ACCAAATGCG AGAATGACAG AGCAGTTTCTT 1200  
TTTCTGTTT TCTTACCCT TTAATCTTTC AGAGTTTAAA GAAATGACG TCATGACAG 1260  
AGACTATGC ATTITGAAC TTGTTGATCC TGGATTTTTT TAAATCATTT TTAATCTAGA 1320  
ACTTAACAA AATATAGAT TGTGTACAGG ACTGTGTGAA AGAAGATGCT TTGCATATTT 1380  
GCTGCATGEC ATCAGTATCT TACTAATAAT GTGAATGAA AGCACTATG TACACTGAAA 1440  
TGCTTAATG TATCTGAAG CACAGGTGA TACTCATTTT TATGTTCTTC CCATTTGTC 1500  
TGTTTTTGCG CTCCTTGACA TCTGTCATCA GTATTTAGAG GTGAGAGCT GAATGTACA 1560  
GGTATAATA ACATTTTAA AACATATAC TTCTCTATTA TCACAGTTCT TCCAGAGCAC 1620  
TCTCAGATAC ATTCTAATGA CCAGACTGG TTTAACAAAA GAAATACAA CCATGGGAAA 1680  
GAAATCTTAA ATGAAACAGC CATCTCATG TACGCAATTT TCCCTCATAT TTTACTGGGC 1740  
CATGTTTGT TCTGTGACT CATGTATTTT TTTTTTCCAG ATCTCTTTCC CCAGTTGCT 1800  
ATTGTAGAG TATTTCTGCT GGTGTGATG CAGTTATACA CATTAAGCA GATCTGGAT 1860  
CTGAAGTAG TATTAAGCAG CTATTAACCA GAATATAGT CATAGCTGCA GAACCATGA 1920  
TAGGTACAGG ACTTTCTCTT TGTGTTGCTT TTTGTTTGT TTTGTTTGT TTTGTTTGA 1980  
CAGAGAGAG ATTGTTATTA CAAGAAAAA AATTCAGTG AATGTGACAG AATCTGCT 2040  
TTTACACCA TCTTAAGAA AACCTTACA AGGGTGTGTT GGAATAGAAA AAGGTTATA 2100  
AAGTTGGAT CTTAATTTGT AATATACCC ATGAGTGTG AAGTTCTTAA AGCAGAGCT 2160  
CATTTGTC ATGAACATA AGGAAGACT ACTGTATAGG TTTTTTTTTT TTTCTCTTTT 2220  
AATGAGAAA AGCTTTGCT TTAGGTTGCG ATACTTTAT TGGATTAAT CTGAATGATC 2280  
CTACTCTTT GGAATTAAC TATGCTTAC CAGTTTCCAA TTGATTTAG CTTCGTGTTG 2340  
GAATTTGAAA AAAAAAGAAA AAAAAAGAAA GAAACCTTA ATAAATAGG TGAAA 2395

55 (2) INFORMATION FOR SEQ ID NO: 160:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2120 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

5 CCCCAGATAC CCGCTGACCT AGTCCCAATC ACACCTCTCG CTTCTCGCG CCTCGAGGC 60  
TAATGAGAC GCTTGGGAAA AGCAGTAAAC GGAATTCGCG GTGACCTTC GCTTTAGGC 120  
TCGTGATTC TTCCGCCAA CCGAGAGGAA CCGGAGAGC AGTTTACAC ACCGCCGTC 180  
GTGTTTAGG CGGCGCCGC TCGCGCGCA TGTTCCTCT TTTCTGTTT TCTCAGAGT 240  
GCTCTGCTTA AGCGGTCCC CCGCACGCAC CATCTGTTTC CATCCCGCC GCGCGAGCA 300  
TTGCGATTT TCGAAGTGG CAAAGTTCTAT GACACCCGTC ATCAGAGACA ACCCTCAGG 360  
CTGGGTCCC TGTGGTTTC CCGACAGTT TCGGATATG CCTTACCAGC GTTTCAGCAA 420  
AGGAGTCG CTAGGAAGG TTGCAAGCTG GACAGAGCC ACATACCAG ATNAGAGTA 480  
CAAAATTAAG TACTCTCTC AGTTGCTGG TGAAGTCAA TATGCTTAT TCCATGAGA 540  
GGATGAAGT AGCTTCGAC TGTGTGATAC AGCGCGACA CAGAGAGCG CCTTACGCG 600  
GAATCCAGT AGATTGCCC AGAGAACCT CCGCAGAGC AAGATCTCT GGAACATGTT 660  
CGAGTTCAAC CTGCAATCC TGCCTAAGAG TCCCAAGAG AAGAGAGAG AACCATTCG 720  
ACTCGAGAA AAGTTCCAGA ACAAATTTGG GTTTAGGCGAG AATGGGATC AGAATACAA 780  
GAAACCCGA GACTCTTCAG TTGAAGTTG TACTGATGG GAAATGAAG AGCAATAGA 840  
TTTTCTCAG TTGATGAGA TCGCTACTT GGAATATCA GAGCCACAGG AATTTAGTG 900  
TTGTGGGCG CTAGATACT AGCAAAAGC GTTTGACCG ATCACAGCA GAGATGAGA 960  
GCCATCGCG ASATNCAGC GCATCTTCCA CACTGTACC ACCACAGAG ACCCTGTAT 1020  
CCGCAAGCTG GCAAAACTC AGGGAAATGT GTTTGCACT CATGCCATCC TGGCCAGCT 1080  
GATGAGCTGT ACCCGCTCAG TGTATTCCTG GGAATTTGTC GTTCAGAGAG TTGGTTCOA 1140  
ACTCTTCTTT GACAGAGAG ACACCTCCA CTTCGACTC CTGACAGTGA GTGAGACTGC 1200  
CAATGAGCC CCTCAGATG AAGGTAAATC CTTCATATCA CCGCGCAACC TGGCCATGA 1260  
GGCACTTAC ATCAACACA ATTTCTCCA GCAGTGTCTG AGAATGGGGA AGGAAGATA 1320  
CACTTCCC ACCCAAGC GTTTGTGGA GCAAGCATG GATNAGATG AATTCGCTC 1380  
TTTTGCTTAC CTTTACCGCA GTGGAAAGCT TCGAGATGAT ATTGACCTTA TTGTCCGTTG 1440  
TGACAGAT GCGTCAATGA CTGAGGCCAA CCGGAGATG TCTTTCTCA ACATCAGAC 1500  
ACTCAATGAG TGGATTTCCA GGCATCTTAA TGGGTTGAC TGGGTTGAG AGCTGGACTC 1560  
TCAGCAGCG GCTGTCAATG CCAAGGAGCT GAAGACAAAC AGCTACAGT TGGCGCGGTG 1620  
GACCTCTGT GCTTTGCTG CTGATCTGA GTACTCAG CTGTGTTATG TGTCTCGTA 1680

CGACGTGAAA GACTCTCTAC GGCACGTCAAT CCTAGGCGACC CAGACGTGCA AACCTTAGA 1740  
GTTTCGCGAC CAGATCCACC TGCACCTGGA GAAATCCCTGG GGCATTTCAC GCTGCGTCAAT 1800  
TGCATCTCTG ATGACCTGG AGAGAGCGCA ATACCTCATC CTGAGGAGCC CGAACAGCA 1860  
GCTATCCGCT GTCTACAGCC TCCCTGATGG CACTCTCAC TCTGATGAG ATGAGAGAGA 1920  
AGAGAGAGAG GAGAGAGAG AGAGAGAGA GAGAGAACT TAAACCATG ATGTGAGACT 1980  
GAGATTTCTC CTTCACCGA GACTACAGAG GCGTTTGATG CTTAGTGAAA TGTGTGTCTA 2040  
ACTTCTCTTC TGCATTTTAC CAGATCAAT AAAATATATA TCTGTTTACT CTTAAGAAAA 2100  
AAAAAAAAAAAAAAAA 2120

20 (2) INFORMATION FOR SEQ ID NO: 161:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(161) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

CGAAGCTGAA GTCTCTCGAG ACCAGAGACA ACCAGAGCAT TCTCTAGAA GCTGACCGA 60  
CCTCCACCTT CACTCTTAC TGCAGACAC AGAGAGCAAA GTTCTCACTC AAACTGATAG 120  
CGAGAGATGG GCGCTGTTC ATGAGAGAGA ACTTCTTCCA GCGAGCGGCC AACCTCTGC 180  
AATTCAGACA GTGAGAGAG CTGTACTGGA CCGCACTGCT GGCATTCCTT ACTCTCATGG 240  
GTTTCGCTGT TGCACAGAC AAAATACAGT TCTTGTGTGT ACCAGACCTG GAGAGAGACC 300  
TTTCACTGGC CTGTGATGTC AACCCAAAGC ATGTGCTGTG CAGAGAGTTC TGTGCTCGAG 360  
GTGACCTGAC GGCCTCTGGA TGCCTGTGAG TTCTCTCATG AGAATAGATA TGTTCATGGA 420  
AATGTGACAG CTGAAAATAT GTTGTGTGAT CAGAGAGACC AGAGTCAAGT GACTTTTGCA 480  
GGCTATGGCT TGCATTTCCG CTATTTGCCA ATGTCCAAAC ACTGTGACTA CTGTGAGAGC 540  
AGCAGAGACC CTGAGAGAGG GAGACTTGAG TTCAATTAGCA TGCACCTGCA CAGAGAGTGC 600  
GGGCGCTGCC GCGCGCGGCA CTTCCAGAGC CTGGGCTACT GCAATCTGAA GTGGCTCTAC 660  
GGGTTTCTGC CATTGACAAA TTGCTCTGCC AAAATATAGG ACATCATGAA CGAAAAAGAG 720  
AAGTTTGTTC ATGAGCGGAG GCGCTTCTGAG GAGACCTGGG GTCACTGATC CAGCGCTCTA 780  
GAGACCTGAC AGAGATACCT GAGAGTGTGAG ATGGCGCTCA GGTATAGAGA GAGCGCGGCC 840  
TTACGCAATGC TACAGAGACA CTGAGAGCT TTGCTCGAGAG ATGTGCTGTG GTTCTCATAT 900

(2) INFORMATION FOR SEQ ID NO: 162:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (162) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

GGCAGAGAT GAGGCGCAC CAGTCTTCT AGGCGACGCT GGGTGTGCT CCGCTAGATA 60  
TGCACCTCTC TTACTTACT CTCCGGGAT GTTAACTTT CTATTTTCA GCTGTGCGAC 120  
CTGTCTAGGC AACCTGACTT CCGCATGGGC CCGTGTGGGT CGACAGAGAC GTGGCTGGCC 180  
CGAGGCGCCA CCGCTCTTT CTGATCTTC TTTCCTTAC ATGATCTTGG GCTTGAGTCT 240  
CGCAGAGAAC CTGTCTTTTA GCTTCACAC CAGAGAGAGA GTTGAATAT ACTTCCCGCC 300  
CGCTCACCA AGGCTGGGAA CAGAGGGAT GTGTGAGAG CAGAGTCTCT GTGGCGCTCT 360  
CGAGGTTTGT TTGCATGAT CACTACTGTC TTCTCTTGT ACTATATCA TCAATATCTT 420  
TCCCTTCCCT GTGGGCACTG GAGAGACTGC TGGGTGAG CTGACACTGC CACTGAGTT 480  
GGGAGAGAG GATTAATGAT GAGCAGCTTT CTGTGAGAG CTCTGATCT AACCCAGCC 540  
CTAGATTTCA GAGTGTGGT AAAGTCCAT GAAACAGAC CTGGCAGCA AACCTGGAAA 600  
TGGCTGAGAG TGGAGAGAGA CTTGACTTTC TCTTCCCTTC TCCCTCTCC AACATTTAG 660  
GAACTCTAC CTGTTAGAT CTTCAGAGCT TGTTTTCCCTG CTGGGTGGGA CAGAGAGAAA 720  
AGAGAGAGAG AGGCTTACGA AGAGCAGCC CTCTCTTCTC CTCTGAGTGA AATGACTTTC 780  
ACCTAGAGTA AATGAGAGGA CGAAAGCCT CTGATTTTGA ATTTCATTA AATGTTAGAA 840  
GTATATATAT ACATATATAT ATTCTTTTAA ATTTTTGAAT CTTTGATATG TCTAAAAATC 900  
CAATCCCTCT GCGCTGAGAC CTGAGTGA GAAGAGAGAA AAATGTGTT TCAATTAAAG 960  
AATGTAATTA AATGATGAA ACTGAAAAA AAAAAAAAAA AAA 1003

50 (2) INFORMATION FOR SEQ ID NO: 163:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(163) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

60 AAGAGCGGC ACAGAGATGT GAGGTCTTAC ACAGAGATGG GAGAGATTAC CAGCGACTTC 60

GGGAAACATC AGCATATGCA TCACCGAGAT GACCTCTATG CTCAGCAGAT GGAACGAGAA 120  
ATGAGGCACA AACTGAAAC AGCCTTAAA ATTTCATTTG AGAAGTAGA GGCCTCTACT 180  
AAGAGGAAAC TGGAAATTTGA AGTGCCTTTT AGGGACTTTGG GATTTAAAGG AGCTCCCTAT 240  
AGGAGTACT GCTCTCTTCA GCGCACTACT AGTGGCTGGG TAAATGCTAC GGAATGGCCA 300  
CCTTTTGGG TCACATTTGA TCAGGTAGAG CTGATCCACT TTGACCGGAT CCAGTTTTCAC 360  
CTGAGNACT TTGATATGAT AATGCTCTAC AAGGACTTACA GCAAGAAAT GACCATGATC 420  
AAGCCATTC CTGTAGCCTC TCTTGACCC ATCAGGAAT GATTGAATTC CTCGACCTG 480  
AATATCAGAG AAGGATTTCA GTTCCCTCAG TCAGCTTAAA TCATGAGAC CATTTGTGAT 540  
GACCTGAGG GCTTCTTGA ACAAAGTGGC TGTCTTTTCC TGGAGCTTCA GGGTCAGGG 600  
AATGATCTG AAGAGGGGA TTGAGAGTCT GAATTTGAG ATGAGACTTT TAACTCTTCA 660  
GAAGATGACT ATGAGAGGA AGAGAGGAC AGTATGAG ATTTATCAT AGAAGGAA 720  
GAGTCAGCT ATTCTTAGGA GTCATTTGGT AGTGAAGAG AGAGTGGAA GGAATGGGAT 780  
GAATCTGAG AAGAGCCCG AAAAAAGGAC CGAGAAATTC GTTAGAGCA AGAAGAGAA 840  
CAAGTCTGA GTATGAGCG GAGAGGAG GCATCTGTGC AGATTTGGG CGTGGCTCT 900  
AACCCTGTT CAGACACAG CTCCTGCACC CCGAGAAA AGAGAGATA ACTTCTGAC 960  
TTTGGCCCTG AGCTCAATTC TTCTCTCAG CAACCCCTCA AATTTTACA TCACATAGAA 1020  
ACTGTATTTT TCCTTTGTG TTCAATTGAA GTTTTGGCAT TTGTGTTTAT GGGTTTAGGG 1080  
GGCAATTTGT GTGACCAAT CTAGCTGGG AATTCAGGC CCACAGAC AGGTCCCAT 1140  
GGCCCATTC AGATGGCAG GAGAGAGTG TTCTTGAGA CAGGAGAG CTCGCCCTGT 1200  
TAAATAAAT TGTTTTCAATC TTCTCTCTTC CTGTCACTT CTGCGAGAC ATTGATGGCT 1260  
TCGTGACATC TATTTGTGT CTCAAGCTG TATTTCCAG ACGTGTGAC AGGTGACCC 1320  
TTAATTACC GTATCATGAT TCTTGACCG CACATTCAT CCTCAACCT ACCCTACTGC 1380  
CATGACCTTC GGCACATTC TAAGTTTAT CTTTGCCATA CTCAGGATC TCGGAATTT 1440  
GCTAATGGT GTGATTAACC ATACAGCTTG AGCCAGTGAG GCAATTTGG CTGTGCTCTT 1500  
GCTCTGAGT TTCTGCTCTT CTTGCTCTGT GCAGATCTG AGATATATCT GCTGCCCTGG 1560  
AAGACATAG AAGCAGTAT ACTGCTGGC TGGGTATTTT TCTCCATACA ATGCAACAT 1620  
GATACATGA TAGAGGCGA AATTCOCAT GTCTCTTTT TTTTCTATA TATCTAAGGA 1680  
AGATATACA GATTGTGCT CATGTACCG TTCTAGTGA ATGTAGAGCA AGGCTCAAG 1740  
GAGTCACAT TTAGATCTG AAGGACAG TCATGCTTG GGCCTAGAT ACCCTGATCA 1800  
GAAAGAGGA GAGGAGGGA GGCATTTCT ACAAACAA CTCTGTGGA CTGCTGCTCC 1860

TTATTTTAC TTGTCTTGC ATTTGCTGT ATTTATACA GTTCTGTGT AACACTTTT 1920  
CAATATTTG GGAATTTAT CTGGCAATC TCCCTTCTG GTTCTCTGCA CCGACTGTG 1980  
CACTGCATC TCTTCCGTG CTCTGTGACT TTAAAGAGAG AAGGGGGAG GGTCTCCGA 2040  
TTTATCTTT GTTCTTTTTT TCTCTTAGC AGTAGGACTT GATATTTTCA ATTTTGGAG 2100  
AATAAAGA TGATAAATC GGGTTTTT TTTGTGTTGT TTTGTAAAA AAAAAAANA 2160  
AAAAAANA AAAAAAANA AAAAAAANA AAAAA 2196

(2) INFORMATION FOR SEQ ID NO: 164:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

GCACAGATC GGGGGACGG AAGGAGAG GGTCTGCGG CGGGCGCTAC 60  
CCAGAGCCA GCGAGCGCA GCAAGAGTG GCTGTGCCC GAGCGCAGCC CCGAGCGAC 120  
CCCCCCCCG CCGCGNAGG AGCGGCTTC CAGCGAGCCC GACTTCTAGG AGGAGGGAG 180  
GGGGAAGGC AGCTCAAGCC TCACCCAGG CCGTGCCTCC AGCCCCGCA CTCGAGGCT 240  
CCTCGGACT CGCGGGTCC TCTCGGAGT CTCGAGGGG AGCGGCTGT CAGAGCGCAT 300  
GGAATGGTG CTGTCTTCC TCTCGAGCT GCTGGCCCC ATGCTCTCTG CCAGTGCAGC 360  
TCGAAAGAG AAGGAATGG ACCCTTTTCA TTATGATTTAC CAGACCTCTA GCAATGGGG 420  
ACTGTGTTT CCTGTGCTC TCTTCTGAT TGGATCTTC CTATCTTNA GTTCAGGTG 480  
CAATGCACT TTCAATCAG AGCCCCGGG CCGAGGAGT GAGGAGGCC AGGTGGAGAA 540  
CCTCATACC GCGAATGCA CAGAGCCCCA GAAGCAGAG AACTGAAGT CAGGCATCAG 600  
GTGAGGCTT CTGAGACTG AGCGGGCTG TTGACCTTT GATGCAAT GTCTGATGCTT 660  
AAGAAACCG GCACTTCAG CACAGCGCTT TTCCCCAGA GAGCCNAGA ACTGTGTGT 720  
CCCCCAGCT ATCCCCCTA ACACCATTC TCCAGCTGAT GATCAACTA ACACCTGCTT 780  
CCCCACTGA GCTTGGCTC CTGCGGCTT CCGGTGATGT GTGTGTGTGT GTGTGTGTGT 840  
GTGAGTGT GTTGTGCTA ACTGTGCTT TTGTGCTTAC TTGTGTGTGT ATGTATTTGT 900  
GTTGTTAGT GAATGTGGA CTCGCTTTC CAGGCAAGG CTGAGCCACA TCGGCATCTG 960  
CTCTCTCTG CCCCCGTGG CTTCTATCAG CTTTGTCTTC TAGGAGGCTG CTTTGTGCCC 1020

GAAGACAGACC CCGTCCCTCTG ATTGAGAGAT GCGTAGAGTA AAGACAGAGG CAGTGTCTT 1080  
CAGTGTCTT GCGACTAGAG AAGCTTTGCA GCACTTTCTC ATCACTCTTC ATTGACTCTT 1140  
TTTCACTCTT TTAACAAAAC CTTCCTCTCT TATCCCACTT GATCCCACTC TGAAGTCTTC 1200  
TTAGACACTG GAGATAGAAA GCAAGACACT GGTAGAGCCA GCGTAGACTT CAGAGAGGCT 1260  
ATGCGCTCTC GTGTATTATT TCTTCCAGAG GCGTCCAGAG AAGAGTCCCC ATTGACCCCG 1320  
CCCTTTGACA GAGCCGCCCGG GATTTCCAGG CCGAGAGGCTT CTAGCTGACC CCTGAGAGAT 1380  
GTGTCCCTCT GATATCTCTT CAGCAATACG TCCATGAGGCT CTGAGACCTT ACCCTTTGCA 1440  
ACCTTCCCTG CTTCAGACAC TTCAATCTAC AGCCAGCTC ATCCAGATGC AATCAAGACAT 1500  
CCCTGCAATT GCGTCTCTCG CAGGCAATAG TTGAGAGACT CCGTCTCCCT TGGAGCCAGAC 1560  
AAGCCGAGAT GAGTGAAGAG AAGACAGAGG CCGTCTCTTC TCTGCCTAGG TCCCTTTTACA 1620  
TGGAGACAG AGGCACTCC CCGATCTCTT GCGTGTCCCG TGGTGTGCA GAGCGGTGAG 1680  
CGAGTGTGAT TGGAGACTCA GCAAGCTCCG TCCAGCCCTT GCGAAGCTG AAGAGTTTGA 1740  
GCTCATACAG AAGTGTGAAA CTGACCCAG AGTCCGCCCC TCCGTCTCTC TCTGTCCCG 1800  
CGAAGACCA CGAAGCCCTG GCGTGTGACC CATTGCTTCT CTCTGTATCG TGAATCTATCC 1860  
TCAACAGCA CAGAAAAAG GATTAAAAAT TCGTTGTCTT CCGTAGTAAA AAAAAAATA 1920  
AAAAAATA AAAAAAATA CTGCA 1945

35 (2) INFORMATION FOR SEQ ID NO: 165:

(1) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 2933 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

GGGTGACCC ACGGTCCCG CAGCTCTCT GTTAGTCTTT GCTTCCCTCG CCGCTCCCG 60  
GATTCAGAG CAGTGTATAC CCGCCGCCCA CCGCTTTGTT GCGGTGAGAG GAAAGAGAAA 120  
GAGAGCCGCC TCCGTTTCTT GCGCGCAGCT CCGCCCACTG CTGAGACAG CAGAGCCCGG 180  
CCGAGTCTGT CCGTCCACAC CCGTGTAGAC CCGTACCCGC GAGCCGCCAC AGCCGCCCGC 240  
CGAGAGAGAG GCGCGCATAG GCTTGTGAG CCGATTCAAA ACGTAGATAC CTATCAGACAG 300  
CGATTCTGAA AAGAGAGAC CCGTCCCAATC GGTATTATGT TGAATAGACC ATCAATAGAG 360  
ACAAGAGTGT GATGTCTCTT TCCAGGCCCA AAGTAGATGA ATTGCACTTG TTCCAGAGTG 420  
AACAAGTGT CCGTAAAGAA AAGAGAGAC GAGAGAGCTT TTGATCTCTC CTTCCTGAGG 480

AATCTCTTC TGAATGAGAG ATTGAGATGA ATGAGAGTGT TCGAATATAC CTTCCTGATC 540  
GCGTAGAGAA TGTATCAGAC ATCCAGCAAT GCGCTGATCT GAGGTAGAGC AAGCGTATCC 600  
AATGCTGACC CATTGATGAC ACGTGTAGAG GCACTTACTG TATATCTCTC GAGGTATAC 660  
TTAGAGCGTA CTTCCTGAAA GCGTATGAC CCAATCGAAA AAGAGACATT TTCTCTTCTC 720  
GTGTGTGAT GCGTCTCTCG GAGTTCAGAG TGGTGTAGAG AAGTCTGACC CCGTATTGCA 780  
TTGTGTCTCC AAGACAGTGC ATCCACTGCG AAGAGAGACC TATCAGAGCA GAGAGTAGAG 840  
AAGAGTCTT GATTGAGATG GGTATGATAG ACAATTGATG TCCAGAGAG CAGCTTACTC 900  
AAGTAAAGAA GATGTGTGAA CTGCCCCGTA GACATCTTCC CCGTCTTAGG GCAATTGTG 960  
TGAAGCTCC TGAAGAGATC CTGCTTTTAC GACCTCTGAG AAGAGAGAG AGCCTGATTC 1020  
CTGAGAGCT AAGCAATGAG ACTGAGAGCT TCTTCTCTCT GATCAATGAT CCGTAGATCA 1080  
TGAAGCAATT GCGTGTGTAG TTGAGAGCA ACCCTTTTGA AGCCTTTGAG GAGCTGTGAA 1140  
AAGATGCTCC TCCATCATC TTCAATTGAG AGCTAGATGC CATGCTCCCC AAAAAAGAAA 1200  
AATCTCATAG CAGATGTGAG CCGGCAATTC TATCAGAGAT TGTAGACTTC ATGAGTGCAC 1260  
TTAGACAGAG GCGACATGTC ATTGTATGAG CAGACAGCAA CAGACCCAG AGCATTGACC 1320  
CAGCTTACAG GCGATTGAT GCGTTTACA GCGAGGTGAA TATGTGAAAT CCGTATCTCA 1380  
CAGAGAGCTT AAGATCTCTT CAGATCCATA CGAAGACAT GAGGTGACA GATGATGTGAG 1440  
ACCTGAGACA GTAGCCATAG AATCTCAGAG GCAATGTGAT GCGTACTTGG CAGCCCTCTG 1500  
CTGAGAGCT GCTCTGAGAG CCAATCCGAA GAAAGAGAT CTGATTGACC TGAAGAGATCA 1560  
GACCACTGAT GCGAGAGCTA TGAATCTCT AGCGATTAAT ATTGATGACT TCCGATGAGC 1620  
CTTGAAGCAG AATTAACCAT CAGCACTGCG GAAAGCCCTG GTAGAGCTCC CAGAGGTAC 1680  
CTGAGAGAG ATCCGAGGCC TGAAGAGAT GAAAGCCCTG CTAGAGAGAG TGGTCTCATTA 1740  
TCCGTGTGAG CAGCCAGACA AATTCCTGAA GTTGTGATG AGCATCTCA AAGAGATCTT 1800  
GTTCATATGA CCGTCTGAGCT GTGAGAAAC TTGTGTGTGAC AAGAGCATTC CTATATGATG 1860  
CGAGCCAGAC TTCAATCTCA TGAAGGCTCC TGAAGCTCTC ACGATGTGAT TTGGAGAGTTC 1920  
TGAAGCCAT GTGAGAGAAA TCTTTTACAA GCGCCGCCAA GCTGCCCCCT GTGTCTATTT 1980  
CTTGTATGAG CTGATTTGCA TTGCAAGACC TCGTGTAGAT AAGATGTAG AGTGTGTGCG 2040  
GCGTCTGAC CAGATCTACA ACGAGATCTT GACAGAGATG GATGTGATGT CCGCAAAAAA 2100  
AATGTGTCT ATCAATGTGCG CTAGCAGAGG GCGTGTAGATC ATTGATCTCG CCAATCTCAG 2160  
ACGTGAGGCT CTGATCTGAC TCAATCAAT CCGCACTTCTT GATGTGAGAT CCGGTGTGTC 2220  
CATCTCAGAG GCTTACTTCC GCAAGTCCCC AGTTCCAGAG GATGTGACTT TGAAGTCTCT 2280

5 GCGTAAATG ACTAATGCT TCTCTGGAGC TGACCTGACA GAGATTTCGC AGCGTGGTGG 2340  
CAAGCTGGCC ATCCCTGAAT CCATCGAGAG TGAGATTAGG CGAGAACGAG AGAGCGAGAC 2400  
AAACCCATCA GCGATGAGG TAGAAGAGGA TGATCGAGTG CCTGAGATCC GTTCGAGATCA 2460  
CTTTGAGAA GCGATGCGCT TTCCGCGCG TTCTGTCTAGT GACATGACA TTCCGAGATTA 2520  
TGAGATGTTT GCGCCAGACC TTCCAGAGAG TCGCGGCTTT GCGAGCTTCA GATTTCCTTC 2580  
AGGAAACGAG GGTGAGAGCT GCGCCAGTCA GCGCAGTGA GCGCGGAGAG GTGCGAGTGT 2640  
ATACAGAGAA GACATGATG ATGACCTGTA TGCGTAMGT GTGTGCGCA GCGTGCAGTG 2700  
AGCTGCGCTG CCTGCGACTT GTTCCCTGGG GTTGGGGGGG GTTGGCCAGG AGAGGGACCA 2760  
GGGTGCGCC CACAGCGCTC TCCATCTCC AGTCTGAA CA GTTCAGCTAC AGTCTGACTC 2820  
TGACAGGGGG GTTCTGTGTT CAAATATACA AAACAAAGCG GATTAATATA AGCGATTTT 2880  
CATTTGGTAA AAAAAAAAAA AAAAAAAAAA CCGGGGGGGG GCGCGAGCA TTT 2933

25

(2) INFORMATION FOR SEQ ID NO: 166:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2243 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

35 TCGAGAGACC GCGCGGGGNG CCGCTCTCGG CCGAGAGCG CCGCTTGAC CCGCTGTAAC 60  
GATGCCGAGA AGTGGCGTTG GCGTGGGAT CACCATAGCT TTCTAGCTA CCGTGTAC 120  
CGAGTTTCTC GTGTATATG GTGTCTATCA GTATACATCC CCGATTTCC TCTATATTC 180  
TTCTTGCTC CCGTGTAT TTCTTGAG AGGGTACG GTGGGAGCA TAGGAGGACA 240  
GTATAGTATG GTGTCTCTG AAAGGCCCA TAGTGATTGA GTCTTCAAAA CCAAGCTTC 300  
TGAGAGCAAG GAGATTTTGG GAAGAAATC TGACTGTGGA TTATGACAAA GATTATCTTT 360  
TTCTTAAAGT AATCTATTA GATGGGCTG ACTGTACAA TGACTCTCG AAAAACTCT 420  
TCACCTAGTC TAGATAGGG AGGTGAGAA TGATGACTTA CCGTGAAGTC TTCCCTTGAC 480  
TGCGCGCACT GCGCGCTGTC TGTGCGCTGG AGCATCTGCG CAGGCTAGG TGGGTTGAG 540  
CAGGTGGGAG GTTCCCAAGT ATTGATTTT ATTCAATGTA TTAAACAG TTCCCATTT 600  
TCAAAGCCTT GACTAGAGC TCATATACA ACCGCAAGT TTCTGTAGT GCGCAAGGA 660  
GTTAGTTGA TGGTGCTTAA CAACATGAA GTATGTGTA ATAGGATTA TATTTATCA 720

60

AAAGATTTT AAAATAGGG CTGTGTTTA AAAAAAAAAA AAACACGAA AACGACAGT 780  
GATTATAGAG AGCTGCACT CTAAAGTGGG TCGCGGCTTC GCGAGCTTC AGGTCAGCG 840  
5 TCGTGGTCC TCGAGTGGG TGTTTACATG GTACACATG TGTGTATCAC CAGTGGGTCA 900  
ACTGCTGTTC ATTCTCCCG TCGCATTTTG TGTAGCAAT CTTACTGAGC AAAGGCAAT 960  
GAAAGCTTT GGTTCACCA CTGCGATATA TTGGAATTTT CAGCTCAGTT TATGAAGTTT 1020  
ATTTCGAAT CCAATGCTAT CTAGAGATCA ATACTGTCT GCGATGTATT TCATCTTAG 1080  
TGAGCAAAA TTGTGTTTT GTTACTACAG AATAGAGATG ACTGTTTTTT GCCACAGCC 1140  
15 TATGGAATTT GCAATCTGTG ATTGCTTTGT AAAAAAGGAA GTGCATATG CACTGCATTA 1200  
AAGCTGTGT GTTCTAGTC AATGATATG GTGAGACAA TGTATCTATT TATGCGCATA 1260  
GACCATACA GACATAATTT GCAAGTATG GGTCTTAAC TTCAAGTGCA ATCTATATGA 1320  
AAACCAATCT GAGCTTGTA TCTCTTAAT ATTTATTTT TTAAAGTGT GAGATGTTCG 1380  
AGAGAGGTT TCCATTCAT TTGAGTGTG CCGGAGGAA ACTGGGCAAT GATTTCTTTC 1440  
25 AGTGTGAAG TTGCTTGTG GTTACAGCTT CCACTGACC CTCAGCTTC GAATACTCC 1500  
AGTTTGTG GTTGTGCTAT TTCTACTTAT AAATTTACCT TTGTGTATT TGCATTTTAC 1560  
AGCTGTGG TTTGTTTAA ATCTGTGAA AGTGGCTTGA TTAAAGACT CCGTTTAAAT 1620  
GGAGCCACC AGTCAGCAGA ATGAGGCTT AGAGGACTT GCGTGTGAGC GCTGTCTTT 1680  
GTGTTGGT TTGTGATGA ACGATCTTG CTGGGGTTTT TTGCTTTGTT TTGAGCGAAA 1740  
35 TGTCTGGAG TAAATTTAA GTTCTGGAG TTAAATTTGT TTACAGGAT TTGTGTTTTT 1800  
AAAAATAG GATCTTCTG AACTTTGGA TCACCCCTTT ATATATTTTC TGAATATGAA 1860  
AACGTTACA TGAATAAAT TTCAATGAA GATGTGACA TTTTATGAAA AACGAGAT 1920  
TATTAGATGA AGCAGCGAG TGATCTTTTA AAACAGACTT GATCAGGAC ACACATTAAG 1980  
TCTTCTCTC GGAACCGGA AGTAATCTTA TATCTTTAG AATTAATGTA GCGAAGGAA 2040  
TGTAAATTT AGATTTTTT TCGCAATAGT TTATAGAAA TATATGACC AAAGTGATTT 2100  
GAGTTTGTA AATGTAAA TAGTATGAC AAATTTTGA CTCTACGACA TTGTGAATC 2160  
TGTGAGCTT CAGTTCTA CTAGTTTTT ACATGTGT TCTTTTGA TTTATTTTTT 2220  
ACTTTTATA AGGTTCAAA ACC 2243

(2) INFORMATION FOR SEQ ID NO: 167:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1816 base pairs  
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

5 GGTGGGAGAC TTGAAATTTT CCTTAAKMG GGGGCTTTAA GGGGAAACTT TCCGGAAATT 60  
TTGGGGTGA CCGACGGCTT GGGGCAAGCT AGGAGAGAA GTTCTTACTT CAGAGAGTGA 120  
GGCAGAGAGC GAGAGTTTCT GGGGGTGAAG GCGGAGACTG AAGTGAAGC GAGAGGAGAA 180  
GAGAGAGTGG GTGGGGGAGA GAGGGGGGCT GGGCCAGAGA GAGAGAGAA ACCCTTCCGA 240  
GAAAGACGA ACGAGCTGAG CTGCTGTGAC AGAGGGAGAC AAGATGGCGG CCGCGAGAGG 300  
GAGCTCTGG GTGAGAGACC AACTGGGGCT CCGCGCGCTG CTGCTGTGCA CGATGGGCTT 360  
GGCGGGAGGT TGGGGAGCG CTGGGGCTGA AGCATTTGAC TCGATCTTGG GTGATAGGCG 420  
GTCTTCCGAC CCGGCTCTGC AGTTGACCTA CCGCTTGCAC ACCTACCTTA AGGAGAGAGA 480  
GTCTTACGCA TGTGAGAGAG GTTGAAGGCT GTTTTCAATT TGTGAGTTTG TGAATGATGG 540  
AATTGACCTA AATGAGACTA AATTGAAATG TGAATCTGCA TGTGAGAGAG GATATTTCCA 600  
ATCTGATGAG GATATGCTTT GGCATTTTGG KTCGCAAT CAGTGGCAT TCGCTGAACT 660  
GAGAGAGAA GAGCTTATCT CCGTGAAGCC AAAAATGAGC CTAGCTTTC GTTAAACTCT 720  
GGTGAAGTCA TTTGAGAGTG AATGATGAGA CTCCGAGAG AGCTTCAATA CCGTCTGATG 780  
GACTTTTAT GTTGAAGCGG ATGAGAGAAA AATGATGATA TTTCAAGCTA ACCCGAGAAA 840  
TCCGAGCTAC GAGAGAGATT TGAAGCGAG AGCGCTACGA AATTGGAGAG GAGAGAGCTT 900  
AAGAGAAATG TCGTCAATAT GCGAAATGAG AATATGAGAA GCGAGAGAGA AATTCTTGA 960  
AGATGAGAA AGTGAAGGCT TTTTAAATG CTCTCTCTCT AACTGTGGGT GAGTTTAAAC 1020  
TTCAGCTCTT GTCTCTCGG TGAATGATAT GCTTGAATAT TGTGTGCA CTGTGTGCTA 1080  
CAGAGCTTGG GAGAGAGTAT AGTTTCCCTC TGAAGAGTGG AATATGATG GTGATTTGGA 1140  
GTTTGAATAT GAGAGAAAGC TAAAGAGATA TCCAGCTCTT TCTCTTGGG TTTTGAATC 1200  
TAAAGCTGAA GATCAATGAG AAGCAGGGCC TCTACCTACA AAGATGATC TTGCTCATTC 1260  
TGAATTTTAA GCAATTTTCT TTTTAAAGAG AAGTGTATTA GAGATCTAAA ATTCAGCTCC 1320  
TCAATGAGCT TTTTAAATGG TTTCAATGGA TATGAGGCTT AAGAAATAC TATTAATATC 1380  
AATTAAGTT ACTCAATCT GTGAGAGAAA AAAAAAGAAA AAAAAAGAC TCGAGGGGGG 1440  
GGCGTTAGC AATCGGCTCT AATGAGATAT GTATTTAT TTTACTATA TCTGATCTA 1500  
TTTGTTTTT KCGTGGGT AATGTTTTT TCCCTTTCT AAGCTATGAG CTGATCTG 1560  
CTCTCTTCA CTTCTGCA TGAATCTGC AATTAAGCTA GTTAAAGAC TGAATATTTA 1620  
GTGAATGGA TCGCTTCTCT GAGAGAGCC CAGAAATCT GTTAATTTGA AATTAGAGG 1680

AATGAGCTT TAAAGAGCT AATTTTGA GAGTGAAT CATTAATAT TATTTGAAT 1740  
AATATCTGC TGAAGAGAAA AAAAAAGAAA AAAAAAGAC RAAATCTGCA GCGCGGCGCC 1800  
GCTACCGAAT TCGCCG 1816

(2) INFORMATION FOR SEQ ID NO: 168:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

20 AAAAAAGCTT GATGAGACTA AGAAAGAGA GTTAAAGCTT CTTTGAGACT TGTGAGAGCT 60  
CAGCTGAGAC CAAGCGAGAC AGTGGAGAC ACGTCAACT TGTCCAGAC AGAGAGAGCC 120  
GAAGCGACA ACGAGGGGCA GATAGAGCTG AAGAAAGATC TACAGCGAGA AATCAAGTT 180  
ATTGAGACTA TCGAGATTT GTTGGAGAG ATGTATTTGA GCTTGGGGAT GATTTTGGCA 240  
TGTGCTCTCT TCTCTCGAAA TTTTACCGAA GTGACTCTTA CAGTGTGGA CTCTGCTTAC 300  
CGATCTGAG GAGCTTTT TTTTATCAT TCTGCTCTCT TATTCATGTC CAGAGAGAAA 360  
AGTTTACGA AGCTTTTGT GATAGAGAGC CTGTTTGA GATATGAG TGTCTGTCT 420  
GCGCTGTGG GTTCAATAT CTGTCTCTCT AAGAGAGCA CTTTAAATCC TGTCTCACTG 480  
CAGTGTGAT TGAAGAGAAA TAAATATCA ACGAGAGATT ATGTTCTTA CTTTATCAT 540  
GATCACTTT ATACAGAGCA CTGCTTACA GCGAAAGCA GTCTGGCTGG AATCTCTCT 600  
CTGATGCTGA TTTGAGCTCT GCTGAATTC TGCCTAGCTG TGTCAATGTC TGTCTCGCG 660  
TGAAGAGAG CTTATCTGA CTTCCTGGG AGTGTACTTT TCTCTCTCA CAGTTCAATT 720  
GATATCTG GATATCTCT AAAAAAGAT GATGAGCTG GATTAAGAGA ACTATGACT 780  
TCTTAAAGAA AAGAGAGAA ATATATATCA GAAAGTGA TCTTAAGATA AATGAGAAA 840  
GTTAAGCTT ATAGAGAGC AAGCTTGA TTTCTTAAT GTTAAGCTTT AAGATATGA 900  
ACATTAAGAA AAGCAATAT TTTCACTGCA TTTAAGAGTA ATGTG 945

(2) INFORMATION FOR SEQ ID NO: 169:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 902 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

```

5  GGCAGAGCCA CAGGAAGAGT GAGGAGACC AGGCTCTGGG GGCCTCTGTG GATCTCTTTT      60
   GTCTCAGAAC TCCGAGCTGC AACTAAATTA ACTGAGGAAA AGTATGACTT GAAAGAGGGG      120
10 CAGAGCCCTGG ATGTGAAATG TGACTACAGG CTAGAGAACT TTGCCAGAGC CCAGAAAGCT      180
   TGGCAGATAA TAGGGAGAGG AGAGATGGCC AGAGCCCTGG CATGCACAGA GAGCCCTTCA      240
   AGAATTTCCC ATCGACTCCA AGTGGGAGG ATCATCTAGT AAGACTACCA TGATCTGGGT      300
   TTACTGGCGG TCCGATGGGT CAACTTCCA GTGGAGATTT CTGGACTGTA TCAATGTGTG      360
15 ATCTACAGC CTCGCAGGA GCTTCACATG CTGTTCGATC GCATCCGCTT GGTGGTGAAC      420
   AAGGGTTTTT CAGGGACCCC TGGCTCTCAAT GAGATTTCTA CCGAGATGT GTRTAGATTT      480
   CCTCTACCA CCACTANGGC CTGTGCGCA CTCTATACCA GCGCCAGAAC TGTGACCCAA      540
   GCTCCACCCA AGTCAACTGC CGATGTCTCC ACTCTGACT CTGAATATCAA CTTTACAAAT      600
25 GTGACAGATA TCATCAGGGT TCGGTGTGTC AACATGTGCA TTCTCTGGGC TGGTGGATTC      660
   CTGACTAAGA GCTGTGCTTT CTCTGTCTGG TTTCCTGTCA CCTGAGGTC ATTGTACCC      720
30 TAGGGCCACG AACCCACAGG AATGTCTCTT GACTTTCGAG CACATCCATC TGGCGATTTT      780
   GCGAGGGGAG GAGGGAGGAG GTAAAGGCGA GGGAGTTAAT AACTGTGTAA      840
35 TCACCCGCTA AAAAAAAAAA AAAAAAAAAA CCAACTCTGG TTTTCAGCTC CATCAGCTCC      900
   TT
40
45
50
55
60

```

(2) INFORMATION FOR SEQ ID NO: 170:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1883 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

```

50 AGAAACACAC TGAANAACCA CATTTTTCTA CATACAGCTG GGGAGTAGC TGAGAACTTG      60
   GCATCGCCA CACATCTAGT GTTGAAAGAG AGTTGAGGAA ACCAGAGGC CAAATGCAATC      120
55 TCGTGGCMAA CCGTGAACTT GTCTCTGGCG CTTCCTCTAC AGTTCTGAG TTGAAATATC      180
   TTTTCATGCC TAGCATCTGC TTGAGTTATA AACCCAGAG CAGCATATGC ATAGACTAGT      240
60 GTTACTCTTT GTTTTGAATT TGTTTTAATG CTTCTTAGA CCCAAGTCCC TCTCTGTT      300

```

```

5  TCTCTCTTGG TGGTAGCCTC TGGGCATCTG GGACCTCAAT CCCAGCTTT CCCACTTTCA      360
   CGAGTCTTTT GCTCTCTTGG CTCTACTCCT AATAGCCCC AGGAGTGGCC TTATGCTTCC      420
   AATATCGAGC ATTTCAAGCT TCTCTCTGGG GATCGGCANT GCGATGGCCA GAATCTGTTT      480
   TCGMTCTGCG GGTATTTTCC AGTGGGTGTA AAGGAGAGCC TGGGCTTTTC CCTCTCTTAT      540
   CCCTGAGGGT GGTATGAGAG GACTGTATCT ACACCTGTTT TCCTCTACTT TCTCTTTTGT      600
   TAGGGAGGCC TCATCTTAGG TTCTCTCAGA GAGTCTCTGG CTTAANGCTG TAGCAGGGT      660
   GTCTAGGTG GGGGATTTGG ACCAANAACG TCGAGTAGGC ATGATFACTG TATGAGGTGG      720
   GCTCTCAAAA TCAGACAGAA ATGGCTTAGG AAGCCGAGG GGAGCATGCC TGTCTCTCAG      780
   TCATAGAGTA TGGGAGGGAC CTCCTTAGCT TCGAANAATG GAATTGAGG GGTATGAGAC      840
   AATAGAGTG CCTAGTTGAG GATGTTGCCA AAGTTTGTGC CAATCTTATC ATTAGTAGAT      900
   TTTATAGCC ACAGACAGAA ACCAGAAAG GAATATGTTT ACTTTGAGTG CTTTATTTTT      960
   TTGTCTTAGG TGTGCTTGG TAGATCAGA AGAATGCTAT ATCTCTCACA TTTTGCCTTT      1020
   AAGTCTTAC GACTTTGCCC ATTTAGTCT AATGGGAGA TACAGATGTG CAAGTCTGCT      1080
   TTTTGTTTT TTGTATTTAT TTTTTTTTTT TTGCTCTG TGATGAGAT TTTGAGACAT      1140
   GCAGAGAGT GGAGAGGATG GTCTTGGAC CCAATGTGTC CATCAGCTAG CTCATCACT      1200
   TATCAGCTAT GGTCAAGCTG GTTTCATCTG TATCTCTCTC TTTTCACTG TATTTGTTAT      1260
   TGAATTTCCA AGACACTATG CCAATCCAAC CCGTACTACT TTGGGAGATT GGTACTCTCT      1320
   TTTGATGCTG ATAGTGTGCG GGTGACTAT CATATCACA TCAAGTCTGC TTTTGTCTTT      1380
   TAATGTTAAC TAATGAAGTT CCAGAGATGG GCTTAGAAA TGTGTTTTAA GAATTACAA      1440
   GGAGTCTCAA AAGAAATGTA GAGGATGCT TCTTTTCCC TTGCATCTAC AAAACAGAGAG      1500
   AGAGACTGTT CTGTGTGAAA ACTCTTTCAA AATTTCTGAT ATGTGAGGT ACTTGAGACC      1560
   CTTACACAGA ATGTCAATCT TTTTTCCTGT GTACATGGA AACTTGTGTG ACCATTAGCA      1620
   TTGTTATCAG CTCTTACTGG TCTCATACT CTGTGTTTGG AAGATAAAT TCGAAATGTT      1680
   TCTGTGTTTC TGTGAAATA ACTTCCCAA AATATTTAGT AACTGTTTGT TCTACTTGCT      1740
   AATTTGACAC CCGTTTANTA AGCCAAATAT TTCTGTGTTT TTAACAGTA TAATAGTTG      1800
   TAAGTTTCCA TGCATGAGG AAAAAAANA ACCTGTATCT CTGTTAANA AAAAAAANA      1860
   AAAAAAAAAA AAAAAAAAAA AAA

```

(2) INFORMATION FOR SEQ ID NO: 171:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2100 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

```
5      TACTTTAGA TTACTGCTT TCGAAAAGTG CATTTCGGA GCAACGTAAA GGTATTCCT
10     TACATCTGTA TGTACACAGG GTACCCAGAG TGTACTGTG GCAAGCTTCA AAAACATACC
15     ATCGAAGAA GTAGGTGTG AGATTAAGAA ACTTTGCCAA ATGAAAGAAA GTGCTACACT
20     TCGAATATCC CCGTCTCAG CCGCTACCGT GAAAGGGGCT GCAACACACT TCCCTGAGCA
25     TCCCTGTGCG ATACAGCTTC TTATATTTTA TATCTTACTG GATGTAGCA TATGTGCTAG
30     GTTCCCTGTA GTCTGCTTCA AGCGAATGTA AGTTTATGCG CATTGAAACA TTTCGAAAA
35     AAAAAAGAT TTAAAGAAAT TAAATAGACC GTAGCTGTGA TTAGATGTG TGTCAATATG
40     GTGTCTTATA AACATAGCAT CCGTGGGTTT AGATGTGTTA AGTGTAGCA CATTCCCTCT
45     CCGTTTGTCT GTACAGGCTTA CATTGAGAAA AATAGAAAA TGTGTGGCTG GCGGATGTGA
50     AGCTCAGGCG GCGAAATGTC GTTCCGAGAT CCGTTAGACA TTACTTTGAC TCCCTAAAAAT
55     AGTAGTAGAT GTATTTGAT GAGTTTGTGT TCCATAGTTC CATTACTGAC AAAACCTGTCA
60     AATAGCTTGA TGTAGAGACA GCAATAGCTA GATGTAGTGA TTCTTACCCA GAGGTGTGAA
65     TAGAGAGAGG TGCATATAAA TAGAGAGAGG TAGACAGTGC ATGATTTGAG GAAAGAGGTT
70     GAAAGAGAGA CATTGATTTCA AAAAAGATGCG TTCTCAATGT GTCTGTGAC TCAACAGACT
75     GCGAAGTTAC ACTTGTGAGG TGTGTCCCTT TCGTTCTAGG TCGATGGCTT CCAATTCAC
80     TTGAATATGC GTCTGTGCG GCAAAAGCAG AATAGCTCAC TTAACTTTTA TCCAGAGAG
85     CTCTGTGTGT CCGTGTGTGTC AATAAAGTGT GTCTTACTTA ACCGAGTTT ACCAAATGGA
90     AATAAAAAGG GACAAACTAT GGAAGATGGA GTGCATGCGA TTGCAGTCCAG CCAAGATTCT
95     CTTTTCGATA TTAGAGACC CATTACATTA GCTACAGGCG AGGTGTGAG AGCTATGTTT
100    CAAATATTTCA AAGATGTGAC GACCTGTGTC TATGTATCAT CATTGATGA ATCCAGTTGA
105    CTCTTTGCGA AAAGGTGTAT ACTTTTCACT AAAAATGCTT ACTGTGCTG TTGATGTTCC
110    TTTCCTGTGT TTACTGTGTC GATTTTCAC ACTATGCAAT TTTTATTTT TTTCAGAAAT
115    CAAATTTTGG CCGTGAAGAA TGAATGAAA AATTTCACTG TAAATGTGTA AGAATGTTGG
120    GATACAGAGA TTTTATTTT CCGTGAAGAC AAATGAGCTG GAAAGAAAC CAACTATGCT
125    TTGCTGTGAG TTTCATCTG ATGATTTTGA CCAATGAGAA TATCTTATG TAAAGGTTTA
130    ATGATGTGTA CAATGTGATA GATTAGAGCG AGATGTGAG AAGCCGAGTT TTCTCTATGC
135    TAAATGTGTC TACTAGAGAC AGCACTTCTT ACTATCTAG CAAATGATA GCGCCAGCTT
140    1440
145    1500
150    1560
155    1620
```

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1930 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 172:

(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

```
5      GATGAGCTGC TATGTGAAAT AACATTCCTT GACTTACGGA AAGCCAGACA AAAACATATTA
10     AAGAAATGT CATTTCATAT ATGTGTGATA GTAGAGAGC CAGGTATATC GTAAAAATTA
15     GGTTCCTCT TTTCCTGTGA TTCTTACAAA ATTGCATTTA TTCTGCAAT TACAAACACT
20     CACTGATGTA TCGAAATGAG CAACTATAGT TCAATGAAA AATAGAGAAA TAAATGTGT
25     TATAGAGAG TGAATTAGAT ATTTCCTTT GTGTATGAG AATAGTGCAC TAAATTTAAA
30     CCGTTTTC TATTTACTT CTATGAGATT TGTCTACAAA TTATGTTTT TCAATGCTCT
35     AATAGATGTA AATAGAAAT TAAATTTCT TTTCGTGTA AAGATGTGA ACTACTTAT
40     TATATTTGTA AATCAATTA ACTTCCTATT ACATTTAAA AAAAAAAAAA AAAACTGAAA
45     CCGTTGATG TGTGCCCGCG TCGAATATG CAGCGCTGCC GCGCGCGCTC GTGTGTGTC
50     CCGCAAGCA GTGTGCGCGG GCGTGCCTCG GAGCTGAGCT GAGCGCGTGA CCGGAGCGCG
55     GGTGTGTG TGCGGCGCGG GCGCAATCG GTGCAGCTTC CTCAAAATTC TGTGTGTAC
60     CCGTTCTGC CAAATGACAC TGAATGATCT GAACTATAT GATGTAGCA AGCTCAGAG
65     AGATTTTGG TTTCAGAGAA AACATGTGG AAGATCAAG GCAAGAGTGS CCAAGAGAG
70     TGTACTGAG TTTCAGCGCA AAGCAATAT CTTTGCCTAC CATTGAGCA TCAATGAGCC
75     TGAATATAT GTGAAATTT TCGAGAGTT TATACTGATT ATGATGCTT TGAATAGAG
80     AGCTGCCGA AACATGTTA AATGATGTG CTTGAGAGCT GATGTGCTTC TTATGTAGAG
85     TCGAACACT GGTATCTTG GACATGTAC TACTATCAA AAGGTGTGA CCGAGTGTTA
90     TGAATGAT CCGTAGCGCA CCGAGAGAG CTTTCTGCG TGTACATTC GTACAGAGC
95     TGTAGTCAAT AATCAATGTA TGTGTGCGC AAAGATCTG TTCAACAGAT TGTTTGAGGA
100    TTGAGAGCT AATCAATGTA TGTGTGCGC AAAGATCTG TTCAACAGAT TGTTTGAGGA
105    AAGAGATGT GATCAAGAG TATCTCTGA CAGAGTGCAC CTTGAGAGCT CTTGAGAGAC
110    AAGAGAGCC GAGGCGAGG CTAGAGATTC TAAAGAGAT GGTGACATTA AACGTATTC
115    AACGAGACC GAGGCGAGG CTAGAGATTC TAAAGAGAT GGTGACATTA AACGTATTC
120    TACTAGAGAA TGGCTTAAAT CAACTGATTA TGAATGAGTT AAATCTTTTA CCAAGCTTTT
125    TAAAGATGAC ATCGAGTATC TGTGACAT GACAAACTTA TGGCGAGAAA GAAACCTGCC
130    1360
135    1420
140    1480
145    1540
150    1600
155    1660
160    1720
```



AGTTCCGTTG GACTGGCTG AGTACNAG TCAGGAGAA GAAACGATG CATCAGTCA 960  
ACAGATGAA CCCCAGTTAG GCTTAAAGA CACAGAGTT CTAGATGTA AGAGCTATG 1020  
ACGTCTTTT TCAGAGCA TCAGACTTT GAGATTTAT TTAGCAGAA AGGGGATGG 1080  
AGCTGAGTC ATATGGATA AGATGAGCC ATCTCGANTG GATTTTGCA CTTCTGCTG 1140  
AAACTCAGG ATGATATTT TCAATATGAA TATGAGAGT AGATTGATA TCANATCAT 1200  
GGCAGGAGC ATTTATCTG CTATTGTAC TACTATGCA GTAAITCTG GATTGATAGT 1260  
ATTGAGGA TTGAGATTT TATCAGGAA ATTAGAGCAG TCAGAGCAA TTTTITTTGA 1320  
TAAACAGCA AACCCAGAA AGAGCTTCT TGTGCTTGT GCACTGGATC CTCCCAACC 1380  
CAATTGTTAT GTATGTCCA GAAAGCAGA GGTGACTGT GGGCTGAATG TCATATAGT 1440  
GACTGTTCT ACCTTACAG ACAGATAGT GAAGAGAAA TTCTGTATGG TAGACACGA 1500  
TGTCCAAAT GAAGATGGA AAGAACAT CTAAATATCT TCCGAGAGG GAGAGAGCA 1560  
ACCTATATAT CACAGAAAT TGTGAAAT TGTGATAGA AATGGAGCC GCTTCAGC 1620  
AGATGATTC CTCAGGACT ATACTTTAT GATCAGATC CTTCATAGT AAGACTAGG 1680  
AAGAGAGTT GAATTTGAG TTGTGTGTA TCCCGGAA AAGTGGGGS CCAACAGCC 1740  
TCAGATGCT GCGAAGCA TACCAATGG CCAATGATGA TGGGAGCTTC AGCCCTCCAC 1800  
CTTCAGCT TCAGGAGC AGATGAGG TTTCATAG TTGATTCGAT TCAGAGAT 1860  
TCTCAATAA TTGCGCGAG TCAATGAG GAAGAGGAG GAGGCGGCC AAGAGGGGA 1920  
TTTAGGNTG 1930

40 (2) INFORMATION FOR SEQ ID NO: 173:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

50 GGGCTTGGCC TCTGGAGTGA GCTTGTGAG GCACTCGGG TGGCTCTAG GGGCAGGAT 60  
AGGGCTGGGG AGCGGGGCC TGTGGCGTG ACCAGCCCT TCTGTCTG GTTCCACCC 120  
GATGCAGTG CTCAGTCT TCAGCGGGA CAGCTACTG AGCACTGCT TCTCCAGCA 180  
CCTCATGTC GTGCTGCT CTCTGAGC CAGCGCTCG CCGAGGCTG TTGACAGGA 240  
CTTCTACTC GAGTTTGG ACGAGACC AGGAGATG GAGAACTAG AGCTATGCA 300  
CTCTAGTGC CTCAGTTTA CTTACCCAG TCAGAGAGAG ATTGGGAGC TGACCTTAC 360

5 TGTGGCCAA AAGATGGCTG AGCCAGAGAA GGGCCAGGC CTCAGCATCC TCTGTACGT 420  
GGAGGCTTC CAGTGGGCA TCCACCCCC TGGGTGCTG AGGGGCCCC TCCGCCCAA 480  
GACATCTCT CTCACAGCT CCGAGATCT CTCTCTGAT GAGGACTGTG TCCACTACCC 540  
ACTGCCGAG TTTCGCCAAG ACCCGGCGA GAGAGACAG TACCGGCTG AGATGGGCG 600  
CGCGTTCGG GACCTGAGC GAGTGTCTAT GGGCTACAG ACCTAGCCG AGCCCTCAC 660  
CTGTCTTCG ATGAGTGTCA AGGTCAITGAC CTCATGGCA GTGTCACTCT GACCACTTT 720  
GGGAGGTGC CAGTGGCCC GGTAGAGCC ACCCAGGGC GTGAAGTCA GTGGCAGTG 780  
TTTGTCCCA GTGCTGAG CAGAGAGAG CTCATCTGC TGTGTGCTG CCACTGGAG 840  
GCCCTGTG GCTGTGAGT GCTGTGAG CTCAGCGCT AGCCCAAGCC ACAGCAGCC 900  
TGTGTGTC AGCTGAGC CTACTGGGC AGGGCAGAG GCTTTGTGT TCTTAAAAA 960  
TGTTTATCC TCCCTTTGCT ACCTTAATTT GACTGTCTC GCAAGAGATG TGACATGTG 1020  
TGTGTGTGT GTTAATCTT TCTATGTTG GAGTONGAA TCCCGGGCC CTCAGGCTG 1080  
TGGGTGCT GTACGCTGC CACAGTGT ACGAGGCTG ACACAGTGT CGTGTCTGT 1140  
GTTGTGGAC GCTTTTAC ACCTGACACT GTGGTGTGA CTTTCTCTC TACAGTCTT 1200  
TTCTGAACT GTGAGTCA GTCTTTTGT GCTGTGCTG TTGCTGTG TGTGTCTGT 1260  
GGCATCTTC TGTATCTCT GAGCTGTGA CAGATGCA CATTTGAAG TCCCACTCCA 1320  
TATGTCTT CAAATGAG GTCTCCCTG ATCCAGAA GTGGAGAGC CCGTGGGGC 1380  
AGGGAGCTG GAGTGGCAG CACAAAGCT GATTCCTGT CCTGTATTC TCTATTCCA 1440  
TAAGCAGAG TTGACAGC TCMAAAAAA AAAAAAAAAA ATTTCTGGG 1500  
CCTCAGGG 1509

45 (2) INFORMATION FOR SEQ ID NO: 174:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3173 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

55 TGAGACCCAS GGTTCGTGC TTTTCCAG AGGTAGAC CTCGAAGAG ATGCTCAGC 60  
ACCACTATG GATCTGTCT CTTTGGCTG AAAGTGGCC GAGAGAGT GGAAGAGT 120  
CAGAAATCT CACAGTGAAT GGTATCTG CAGAGTCACT CACTTTCTCT GTAAATATCC 180



5 COACATGGA GACTCTGAG ATCATTAAGC CCTTAGATGT GTGCTGGTGG ACCAAGAAC 300  
TCTGTGGGTT CTAGTGGAC AGGGTGTTC AGGATCATCA GGAGCCAAAC CCCAAATCT 360  
TGAGAAAT CAGCAGCATT GCGACTCTT TCTCTACAT GCGAANAAT CTGGCGAAT 420  
GTACGAAAC GAGGAGTGT CACTCAGGC AGGAAGCAC CAATGCCAC AGATCATCC 480  
10 ATGACACTA TGTATGAGTG GAGGTCCAG CTCTGGCAT TAAATCCCTG GGAGAGCTG 540  
ACGTCTTCTT ACCCTGANT ATTAGAATC ATGAAATAT GTCTGAGT TGAATGAG 600  
GAACCTGTAT AGTATCCAG GATGAACAC CCCCTGTGGC GTTACTGTG GGAGACACC 660  
15 CACTTGAAG GGAAGGAGA TGGGAAGGC CCTTTCAGC TGAAGTCCC ACTGGCTGC 720  
CTCAGGCTGT CTATTTCCG TTGAATATG CCAAAAGTC TACTGTGTTA TTTGTATTA 780  
20 ACTATATCT CTGAAGGCG CTGAGGCCA TCTGGGAGT AAAGGCTGCG CTTCGCATCT 840  
AATTTATGT GAGTCAAT ATCTCATGTC TGTGATGCA GCGAAGTAT ATCTGTAGT 900  
ACACATGTA CTGATGTGTT TTTCTGAAT ATTTGCATAT TTTACTTAAA AAAAAA 960  
25 AAAAACTGA GGGGGGGCC GTACCCATTT T 991

30 (2) INFORMATION FOR SEQ ID NO: 176:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

40 ACGCCCTCT TGGAGGCTG AGCGGGCTC TCTCTACTA CTTCAACCC CAGGGGGCC 60  
CTCAGAGGG CCCCTCTCT GCTTGAAGG CTCTGCTGTT CTCCCGTCC CCTTGAGAG 120  
45 AACAGGCCA TGGGTGGCC CTCTGCTG CTCCCTCTG YCTGTCTG GCGCGACGA 180  
TTCTGCGGC CTGTGCTCT CACAGATCT GGTCAAGCT ACTTTATG GGTACTCAA 240  
50 CAAAACACC TCTAGGCTC CATGGTGGC TCTGTGAAA TCCCTCTCT CTTCATATAC 300  
CCCTGGAGT TAGCAVAG TCCCTAGTG AGAATATCT GGAGAGGGG CCATCTCAC 360  
GGGAGTCTT TGTACAGAC AAGCGGCTT TCAATTCACA AGAATATGT GAACGGGCTC 420  
55 TTTCTGAGT GGAGAGGGG TGAAGGAGC GGTCTCTCA GATCTCAA CCTTGGGAG 480  
GAGGAGCAT CTGTGTATT CTCCGATC GAGCTGACA CCGAGATC AGGAGGAG 540  
60 CAGTGGAGT CATTAGGG GACCAATCT ACATCACCC AGGCTGTAC AACCAACC 600

5 ACCTGAGGC CGACAGCAC AACACATA CCGGGCTCA GGGTCACGA AAGCAAGGG 660  
CACTAGAT CATGCCACT AGCTCTGAC ACTGCCATCA GGGTTCATT GGTGTGGCT 720  
GTGCTGAAA CTGTCATTTT GGGACTGTG TGGCTCTCC TCTGTGGTG AGGAGAGGA 780  
AAGTAGCAG GGCGCAAGC AGTGACTTCT GACCAACGA GTGTGGGGG AAGGATGTG 840  
10 TATTAAGCCC GGAGGACGTG ATGTGAGAC CCGTGTGTAG TCTTCCACAC TGGTTCCCA 900  
TTGGCAAT ACATGGAGAG CACCTGAGG ACCTTTAAA GGCANAAGCG CAAGGCAGAA 960  
GGAGGCTGG TCCCTGATC ACGACTGCA GGAGAGTTAC CTACAGAGC CTTATCTCAG 1020  
15 GAGCATCAC ACTGCATCA TATAGGAATG AGCTGTGAC TCCACTGAT TAAACCACTG 1080  
GCAATTTGGG GCTGTTTAT ATAGAGTCC AAAGATTTCC TTTATCTCC CCAAGGATG 1140  
AATATACAT TTTTGTGCT TACATACAC CCGTTTCTC CTGTGCACA TTTTCCATC 1200  
20 TGTATGTTG CTGTCTCTA TGGCAGAGG TTTTGGGAA TAAATAGCT GAAATGATC 1260  
TGACTTAAA AAAAAAANA AAAAACTCA 1290

(2) INFORMATION FOR SEQ ID NO: 177:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

60 TGGGGCCCT TTTGGATCT CTGGGTGTT TTGCGAGAG TTACAGGATG TCAAGTGTG 60  
40 GGAGCTGAC ACCCTGCTG TGGACAGTG AAGGCTGTT CAGACAGGT GCTTCCAGAC 120  
ATTTCAGGC TCCAGGAGG AGGCTGGAG CCCCACAGA AAGCAGGA AATGCAAAA 180  
AAAAAAGT CTTTTTTTT TTTTCTTT TTATTATGA AACAAAACAA ATCCCCAGG 240  
45 AGAGGCTCC ATGATACCA GAACATCA AGAGTACTTT CTACATTTT TATTTCTGTG 300  
TGTTCAGGC AGCATGCA TAACAGCT AACTACTTA CATTGACTC ATTTTCACTA 360  
50 ACTGATTT ACAGATAT ACTAGAACG GCATTAATA GTTTAAGAA AGTTAGGTA 420  
AATTCATG CACTCATAC AGAAAGTAA CATTTAAT ATAAAAAGA AAACTTCTT 480  
GGAGCATTA TCCAGTATT AAGGACAGT GCTACTCTG ATGTGACAA TTTCTATGT 540  
55 GGTGTTACT CTTTCCAAA AGCTGTGAG AGGGTGAGT GCTGCAAG CACATACAA 600  
AAAAAACA CAAAAAAA TGTGCTTAC AGTTTGTAG CAAGATGACA CTGCCAACA 660  
60 CAAGAGGGG TCTGGAGTTC ATTTACGCC GGAAGCTGCG CCCCCTGGCC TCCAGGGGTC 720

ATTGAGAGTG TTCTGAAATC GAAATTCGAC AGAGAGCTTG TCACTACTCC TCTCCCTCTG 780  
AAAAAGCAT GTTGAAGCT GCGGTACAGG TCTGAGCAT GAGGAACTT AATTGAATCA 840  
GCGAGCCCTT CTAAATACAG AGAAAGGAG GTGACTTCA GCTTCAGCCC GCGAGGAGG 900  
GCGTGAAGA ACTGATTAAT CCTTCGAAC TCTGAAGAAA AACCACTTT GAACTGTGA 960  
GAGAAAGCT TAAAGAGTG AGAGACCA TCGAGCTGT GAGTGAACC AGCGTGGCT 1020  
GTGAGGTCA GTGCTCTCT TCTAATGAA AGAGAGAGAG GCGAGCCGAC GCGAGCTTCA 1080  
GAGAGACCA GCGAGAGAG AAGAGCCCA GAACTCTCC TGTGCTGTC GCGAGAGCC 1140  
GCGCGCTCCC GCGAGGCTC AGAGAGGAG CAGCTCAGTG CCGTGTGGCT GAGAGGGCAT 1200  
TGCAGGAGC GCGCCCGAGC CCGAGGAGC CCGGACTTGG GGTGTAGCTA TCAAGCCAGC 1260  
CTGTGCTGAC AGCAGCTTAC GAGAGAGCT GCGTGAAGAC CTGTCAACTG TCGTGTGTA 1320  
ATTCTTAAA TTGGTTTAA ATTGTCATTT AAGATCTGT TTAGAAATTA CCTTGTGAAA 1380  
GAGAGGTAC TTAAAAAT GGAACCTTTC AATTCATTT ATATTTTAT TTATTAACAA 1440  
ACTTAATTA AAGTTTACA AACTGCTGA AACTCAGCA AGTGTGAGC TCAAGACCA 1500  
TTTAAAAAT GATTAATTAC CAGACTTCC TCACTAGAT TCCCTTCCA GTAAAGTAT 1560  
AAGTAAGCT GTAAAGGTCA GTGACTCTG AATCAATTT ATGATTTT TTAAATCAC 1620  
GTGATTTAG ATACTAATGA TATGCTCTAT ATCTATCAG AATCTGTGC AGAAAGCAT 1680  
GCGACATTA GAGAGAGAC CAGACAGAG CTCCATNACC ACGTCTGCC TGGAGGCTCC 1740  
GCGAGCTGA GGTCCGGAG AATGCTGCT TTCACTCAT TTCCGACTTA CTGTGACAC 1800  
GCGTGAAGG GAGCAGACT GCGAGTCTCC GAGAGGGAAT CCTCTTGGGG CCGAGAGCT 1860  
CCTCAAGCC TGGAGAGGC AGAGAGCTC GAGAGGCTT GCGCAGGCA CTGAGAGCTG 1920  
GAGAGAGCA GCGATTTCA CCGGAGAGC TGGAGGCTA ACTGTGCAAT TCTTGGCCGG 1980  
AAGCGCATG CTGATTTGTT GCGCATTTTG GAGATCCCC GTACTCAAG AGCAATGCC 2040  
AAGCTGTGG AAAACAAAC CAAAGATCA CTTCTATTA AACTGTAT ATATATATT 2100  
TTTAAATNG AAGTTTAAA ATCAAGACT AGATTACTA TACATTTT CTCTGATTT 2160  
AGAAATTTA TATTATTTA CTGGGCTCC CTAAATGAT TTCTTTTAA ACAGCTTTA 2220  
AGAGACCA AGTAATACA AAGAGACTA AGAAATTAA AATTAAGAT GTCTGTGAC 2280  
TGAAGCTGT 2290

60

(2) INFORMATION FOR SEQ ID NO: 178:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 549 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 178:

GCGAGAGC ATGCTGGGC TCTCTGAT TCTTCAAGTC ACTGTGTG CTGTGTGTA 60  
CTGAGAGCT AACTGCAATG TGGCGAAG ATGACCAATTT CTTTCCAG AGGCGAAAA 120  
TGTGAATNG TGTCTTCCA TGCCTCTT CATGGCTTAC CAGCTTCCC AGCGTGTTA 180  
ATCAATACA ACCAGAGAG AAGCTGCTG AACTGACTTC TGGAAATGCC CTGGATGCT 240  
TTGTGAGAG AATGTAGTNG GCAATACAT GATTGCTGAG ATCTGGGCCC TCTTGAATG 300  
AGTGAAGAG TAAAGGCCA CCAATCTCTT GACTCTTGG GAACTGATCC ACAAGAGA 360  
TGTTCGAG ATGCTGTGTA AGATTGCTA AAATTAAGG GTTTCAGACC CCGTAATGC 420  
ATCAATCTA GAACTCTCT TCAAGAGAG CAGAAAGTC ATTACAGAA GTCACTGTA 480  
AAGATTCAT CCGAGATTT GAGTGAATC AATTAATTT TCTAATTTT AAAAGAAAA 540  
AAAAAAA 549

(2) INFORMATION FOR SEQ ID NO: 179:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 179:

GCGAGAGG CTGATTCATTT CCGAGCCGGG CCTGCGAGC ACTGAGCCC TTCTGAGACC 60  
GCGCGCCGGA TCGGACCCCG CAGCCCGGAG CATGTGAGGC CAGAGCTGC AAGAGCGTTC 120  
GCGATTCAG ATCTGCGGGA TCAATGGGTC AAGTAATGC AACTGTGCGG GCAATGTCA 180  
GCGGGGAGAC ATCTGAGAA TGAATGAGA GCGAGAGGCC ATGATCAGCA CCGGCAATG 240  
CAGAGCCAG AACGGGAGAC GCTGTGTGC CAGCTGTGCT CTGTGAGAC GAGCGAATTT 300  
CCTGTCTCC ATGTGCAATG GTGAGGTGAC GATGTGAGC GCGAGATTA CTTACACTC 360  
CAGCACTCT GTGAGGTGC AAGTCAAGT GATGTGAAA AAGATCTCA CAGTGTGAAA 420  
AAGCTGACC AATTAAGCCA CCGTGTGTA TGTGCCCCG TCGTGAAGA ATGTGACAAA 480  
GCTCTGAG GTGCTCTCTG TTGTGATTC CCGAGAGAG CAGAGAGAG AGGCGCGAAA 540  
GCGTATGAA GCGAGAGAC TGGAGGCAAT GAGAGAGAG TGGAGAGAC GAGATGCT 600

409

5 CGAGCCAGTC CTCACCCAG AGCCAGACAC TGTCACTAC AGCCAGTCCA GCTTGATCCA 660  
CTGTGTGGG CTTTCAGACT GACCTTCCA GGGTTTGTG CAGGAGGTG TCACATGAA 720  
GCTCATGAT GAGTGTGGG GGTGTGTGG TCACAGCCAC TCCAGACCA ACATGTCCAC 780  
AGCTTCCGTG GACCCCATTA ATTTCATCA CAGATCAGA AAGGCTGGG TCATCAACAT 840  
10 CTCGGAGCC ATGACCTTCA CGAGCAATTA GTCCATGGAG ATCGAGTGT TGTGTAGCC 900  
CGACCTGTT GTGACAGCT CTCAGAGCG CTACCGGGCC GCGATGTCT TCTTCACTTA 960  
CTGTGTGGT AGCCAGAGG GCAAGTGGCT GCTGTGGCC CAGCTGTGG CCGAGAGCCA 1020  
15 GAGCAGAGG AAGGCTTTG AGGAGGCCA AGGGGGTAC CTCAGATCA AGCGAGGCR 1080  
ACAGGGCCAC GCGGAGTTC AGCCCTAGAC TCCCTGTCC TCCACTGTGT GCTCTGATTA 1140  
20 GGCATGCCA GGGGCCAGT GTCCAGTAC TTAGAGTTC GCGCTTTGG CAAAGACCCA 1200  
ATTCAATTTG AGAGTGGTG TTGTGTGAG TTTTGTATC ACGATGTTAA CTTGTACTCT 1260  
25 CTCCTGCMA CTCACAGCC AAGCTTTAT TTATATCAT CAGTATGAA TGTACACAG 1320  
TGTGTGGG AGCGCCGGA GGGTTGGC AGAAGCCTC GGGATGCTT CCGAGCAGCC 1380  
TTAGAGTAT GGGAGAAC CAGCAGACT AATAAGCTG CTGCTTGGCT GCAAAAAA 1440  
30 AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1500  
AGAAAAAA 1509

35

(2) INFORMATION FOR SEQ ID NO: 180:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1316 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

45 AGCTGTATCA TAGGAAGAT GGCACAGG GGGTACAG TAACTCTCC TCCGGCCAG 60  
CGAAGCCAG TCCGGGGGC GGGCCAGCC TCAGTTCCAG GGCAGAGCC AGCAGCGCT 120  
GGGCTCGG TTCCGCTGC GGTCTAGCC TGCATCTCA GACCTTGGG CAGCAGGCC 180  
TCCAGTGG GCTGTGGC AGAGCCGGC CTCAGGCCA TTCAGGCCA GACCCAGCC 240  
55 CCGCTCTGC CTGTCTCTC TCTTCCAGG CCGTTCCCG GCGCCCGGT GTCAGGGTG 300  
CAGCCAGTCA TTTTGGCTC CATTTGGAC AGCTAGAGA GAGCCAGCA GGTGCTGCC 360  
CGATTTATG GAGCCTGTT GGGAGTCTC GAGAAAGCT CAGTGGAGT CACCAATTC 420

60

410

5 TTTTCACTCC CCGCAATGA CTCAGAGAT GAAGTGGCTG TTGACATGA ATTTCCTAG 480  
AATATGTAG AGTGCATTA AAGGTTTCT CCAATAGAC TCATCTCTGG CTGTACGCT 540  
ACGGGCCAG ACATCAGCA GCACTCTGTG CTGATTCAT GAGTACTACA GCGAGAGGC 600  
CCCAACCCC ATCCACCTCA CTGTGACAC AAGTCTCCAG AACGGCCCA TCAGCATCAA 660  
10 AGCTTAGCT AGCACTTAA TGGAGTCCC TGGAGGACC ATGGGAGTGA TTTTCAAGCC 720  
TCTGACAGT AATACCGCT ACTAGACAC TGAACGATC GAGTTGACC TCATCATGAA 780  
GACCTGCTT AGCCCAACA GAGTGTGG ACTCTCAGT GACTTCCAGC AAGTACGAG 840  
15 GGCATCAGT CCGATCCAG ATGCCCTGAG TACAGTGTG CAATATCCAG AGCATGACT 900  
GTCTGGAG GTCTCAGCT ACATACTGT GGGCCGCTTC CTGATGAGC TGGTTAACCA 960  
AGTACCGAA ATAGTTCGG ATGACTTTGA GACCATCTC AACACCA CAATAGCTCT 1020  
20 TTTGATGCT ACCTACTCG CCAACTCAC AAGTCCAG ATTGCCACT ATCAAAACT 1080  
TGTAAACTG TGAATGACC CCAAGCATA CACTTCTCTG TCTAGATATT AACCCAGGA 1140  
25 CTCAGAGT AAGGAGAAAT GGGTTTTTG TGTCTTGG TCACTCTGAG ATAGTCACTT 1200  
GTGTGACT CTAATAAAG GAGCTTACT TTTGTAAAT AAAAAAAA AAAAAAACC 1260  
30 SERGGGGG CCGGTCCA TTSSCCCTT NTAAATCT NTAAATCT CCGGCC 1316

(2) INFORMATION FOR SEQ ID NO: 181:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

45 GGCATGACA GACATGACT CTATTCGAG GCTGTCAAG TGGCAGGCTC ATGAGGAGA 60  
CATCATAG GGTGCTCTT ATCTCTCTG CTCTGGATC CACACGGGG CTATCTCGT 120  
TTATGGGCT GGGACTAGA ATTGCATCT TCAAAACCAT CACCTCTTGG CCAACAGTT 180  
TGACCCAG GTAGATGATA ATCTCTTCA GTCTTAGAA GAATACCTAC GTTATAGGG 240  
CAATCTATT GGGACTGAA CTTTGAAGC CACATATTG AAGAGGCTT GCTTACCTGT 300  
TGGGGCCAA GAGGCACTT ACCAAACAT GTTCAGAA VTTGGYKGG AICARKEKGG 360  
55 GKGGGARRH CRGGGYTGG SCAMTTSK KGGGMCCTT TTAGGTAAR RRGCGKGTW 420  
ATTAGTGT GGTAAAGTA GATCTTTTG CCGTTGCAA TTTGCTGCTT GGTGTAATG 480  
60 TGTGTGTG TTTCTACCC CTAACTCTAG TAGTCTCTC ACTACTTTC TCATTAAGTG 540

AGAAATGAGA CTGCTGTGAT AAGAGAGACT ATGTGTGAGA GCACTGTGAT 600  
TCACTGTGAT GCTGCTGTGT AGCTTTTCA TTCTGTGAG CTGCGCTTCC TAAATATCTCC 660  
AGCTTTGTGA GCTGTAGAGA GAACCTGTAT GAAATAGAGA CCTTCCCTTC TGTACTGTGA 720  
ACTTAAATAT AATATGCTCC TGATTCAGAG TAAATATATAT AAAAAAAAA AAAAAA 777

(2) INFORMATION FOR SEQ ID NO: 182:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

GGACAGATTA ACTATGTAGA TGTATCTCTT AATATTTT TTAATTTT TATCTTGGC 60  
ACTGTCTTTC AATGTGTAC ATGTGTCCA GGAAGAAAT GCTTCTGT TTCTAAATT 120  
GATCTTTAG ACTGTCTT TTCTCCATCT TCTAGATCCC TGGCTTCT TGAAGGTACT 180  
TCCCTGGCCA GAACCCCTCC AGGTCAAGAG CAGAAAGAAA GCTTCATGAG TCAACAGAC 240  
AAATGTGGGC TGAATATCTA TGTATTTGAT TTTGACTTGA ATTTCGTGA TGTTTTACT 300  
GATCTTGAGA AAGAAATATC TTTCAGAGAA AATATATATC CTTTGAAT GTTAAATAT 360  
CCTTCAGCAC CTTCAGAGAC ATTTCATATA TGTTTTCTCA GCAAAATGAG TTATATATCA 420  
TTTGTGCTCT GAGGTAGAG TCTCTAGAAA TCAAGACACT TCACTGATTA AATCTGACT 480  
TGGCTCTGGA CTGAGCTCTG TGAAGTGGC GTTCTGACT GTCAAGATA GGTGTAGACA 540  
GACGACACT ACAAGAGCA CCTGTGTGGC TGTGTACTAG GAGCCAGACT CCTTCAGAGC 600  
GCACTAATAG AATTCAGACT TTACATGAG CTGTTCATTA CTGTTCGAT GAACCTGAT 660  
TGTGTCTTCT CCTTCAGAC CCTTCATCAT TTCACTGTTT TGTTTAGTT TTTCTGTCTT 720  
TTATATATTA CTACAAATTA AATCTTAG GCTATCACC TTACTGTAT AAAAAAAAA 780  
AAAAAAAACT C 791

(2) INFORMATION FOR SEQ ID NO: 183:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

AATATGATTA AGACTGTGA AAGAGCTCT GATTGTGAG GCTAAATAG CAGACTCTCC 60  
AATAGCATAG AACTGTAGAG ACTTGTGAT GAGAAAGAGA TCGAAGAGGA GAAATATGAG 120  
AAGCTGATAG GCGAATATCA TCACTGTAGA TCGAATTTAC AGCTATATCA GGCACAGCA 180  
GTATATGAG CAGATTCAGC AAGAAACAG TTACAGATTC TCGACAGCA AATAGCTGAG 240  
CAGAAACAT CCAATAGAGA ACTATAGACA GAATCTGAG GATTCAGCA GAAATCTCAC 300  
TATATAGAG AAGATCTTTA TCGAACAGAG AACATCTCC AAGCAATAT TAAATATCCA 360  
GACAGAGAAA TTCAAAATCT CAGATTCAGC CTACCAATTA AATCTTAG CAAATCTCAT 420  
CAATCTGAT TGAATATGAG ACTTCATGAG CTACAGAGTA CTCTCATCCA GAACAGAC 480  
ATCTGTGAGA GTTCAGACAC AAGAAAGAC TCCCTGTCT TGCACGTGA GCGCTCTGAA 540  
CAGAGATTA ACTGCGCTTC TGAATGAT AGTAATGAT CTTCATTTA TATGTCTGGA 600  
ATTAGCATG GTTAAAGCAC TGTCTGTGA AATGTCTG TCTTTTAA TGAACAGAAA 660  
ACTATATGAG CAGAATATTA CCAAAATTT CCAAACTAGC CTATCTCAT TGAATCAATTT 720  
AGTATTCGC TGGCAATTTT TCTTCAGAA TACCCATAG CCGAGTTTT TGTATATATA 780  
TATATGCTT TCTTCTACT CTGAGTCATG ATTGTCTGT TCACTTACAC ACCAATAG 840  
CAGACAGAC AACATATGAG CAATATGAC AACCCATGTT GTTCATGTA TTTGTCTCT 900  
TTTTCATAC TTGCATCTG CAGAAAGCC AATGCTTAA AATTTCTGAG AACATCTCAC 960  
AAGATATTT TATCATATCA AACTTTTAC TTTTATGAT ATTATAGAG TATTTTACT 1020  
AATCTTCCA ATTCTTTA AATGTATGA GTTCTTAAA CAGACATTA TTATACAGC 1080  
TCACTCTGAC TTATCTGAG TTATATGTC CTAAATATTA TGTAGAGAAA GATGTGAGG 1140  
TTGTCTACT CTATACAGC CACTGTATG TTAGGTGACA TGTATATGAG GTTATATCA 1200  
GGAATCTTA TTGTATGAG TGCAGAAAT AAAAAATTT TTTTATATA TTCACTCTCC 1260  
TTTGAATTT TCAATATAT AACTTTGCA AAGATTTAG TTGTATAGAG TTACAGCTT 1320  
GATGTGTGA AATCTTTTA TAAATATGA AATTAAGTA AATATCAT TTTCTTTTC 1380  
TAAATATATA AAAAAAAAA CTCCA 1405

(2) INFORMATION FOR SEQ ID NO: 184:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1596 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

413

414

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

5 GTATGAGAT GCGCGGAGCA ACTGTCTCT TTGAGGCTCGA CCTTAGGGG CCGAGAGGA 60  
AACTGCGAG CGAGGTGAC CGGGAGCGA GCAITTCAGA TCTCTCTGGT AGACTGTGTG 120  
CAACCAACC ATGTGGCTG CAAGCTGGT GTGTCTCGG ACATCACTT CTAGGTTT 180  
CGACCAAGT TTACCAAGG CTTCCCTCT TGTGAAGAT TCCATCAGA AGAATCAATG 240  
GCTGTACCA CTAGCAGGG AATATGCCAG CAAACAGGA ATGGGATCC GCGCTGGAG 300  
AACTGGCAA GAATCAAG AGCCAGAT GAAACATCG ATGGAAATA TATTTAAT 360  
TGAATAGTG GGAATAGT TTTTGTCTG AGGGCTGCT GTTGTCTTG GAGCATTTG 420  
CTACTATGC TTGGACTGT CTAATGAGAT TCGAGCTATT GAAAGGCTG TAATTTGGC 480  
TCAATATGC AAGATAGAA TTCAATCCAG CTATATGAC TTAGCAGGA GTATGTTTT 540  
AAGAGTTTG TCTCCATAG CAATCAGAG AAGGCTGTT CTCATGAAT TCAATGATG 600  
AGGCTCTTG GTACAAATG GTTGACTTT TCGAGCATG GTTGAGCTG GAATCTGAT 660  
ACGATCATA CCATATGACC AGAGCCAGG CCCAAGCAT CTTCCTGTT TGTACATTC 720  
TGTGTGATG GTTCAGTGG TGGCTCTCT GACAAATTA GGGGCTCTC TTCTCATCG 780  
AGCTGATGG TACAGAGTG GATATGGG AGGCTCTCC ACTGTGCCA TGTGTGGCC 840  
CACTGAAAG TTTCAGACA TGGGTGACC CTTGGAGTG GGCCTGGCT TGTCTTTGT 900  
GTCTCATTG GATCTATGT TTCTTCACC TACCACTGG GTGTGTCCA CTCCTTACT 960  
AGTGGCAATG TACGCTGAT TAGTTCTTT CAGCATGTC CTTCGTATG ATACCGA 1020  
AGTAATCAAG CTTGCAAG TATCAGCAT GTATGAGTT CAAAAATATG ATCCCATTA 1080  
CTCGATCTG AGTATCTACA TCGATGAT AAATATATT ATCGAGTTG CAATATGCT 1140  
GGCACTGGA GCGACAGAA AGAATGAG TCACTAGCT TCTGGCTCT CTCTACATC 1200  
AAATATCTG TTATATGGG CAGATGCA TTAATATGT TGTACAGCA GCTTTCTTG 1260  
AAGTTTAAA GATAAGAAC ATGTATCAT ATTTAAATGT TCGGTAAATG TGAATGCTA 1320  
GGTGTGCTT TTTTCTGGA GAATAATGC AGTAATCTC TCCAAATATA GCAACACAT 1380  
TTTCAATCT CATCTTGAG TGAATTTAAA ATGTTTTGAT GAATGTGAA ACTAAAGTTT 1440  
GTGTATGAG AATGTAGTC TTTTCTTAC TTTHAAATTT AGTAGTTTCA CTGATTAAT 1500  
AAATTTAGC AAACCTGAT TTGCATATT TTTGAGATG CAGATATG TAATTAAGC 1560  
ATTCCAGTA NAGTGTNTT AAGTGTGTC TATATN 1596

60

(2) INFORMATION FOR SEQ ID NO: 185:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2293 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

5 GCGAGAGCC CTAAGAGCA GAGCAGCAG GACAAGGCG ACTCAGGCA AAGCGGCTG 60  
ACCTGATGG AGGAATGCT CTTCTGGGC CTCAGGACC GCGAGGTTA CACATCATTT 120  
TGGATGACT GTATATCATC TGGATTAAGT GCGTGTATGT TATTTGATTT AGCATTTGGA 180  
GGAAGTTAC AACTAGAGCC TTGTGGAATG AGAGTAATA GTCTATTAAAC AAGAAAGGTA 240  
ATCTGTAAAT CAGATGCTCC AACAGGGAAT GTTCTTCTTG ATGAGCTCT GAAGCATGTT 300  
AAGGAATCT AGCCTCCAGA AAGGTCCAG AACTGATGG AATTACTTNG TCGTGAACA 360  
TGAATCCAT TAAATTTGCA TTATCAGTAA AGAATATPAC GCGAAGCATT AGCTAAATAC 420  
CTGTGTGAAA AGGATGATTT GACACAGAG AACAGAACT TCTTACTTTT TGACATGACA 480  
ACACATCCC TCACCAATTA CAAATTTAG CAGCGCTCA TGAAGAAAT AGAGGAGGCC 540  
GTCTTTGACA AATGGGTGAA TGACCTCAC CCGATGACA GCGCTTCTT GCGCTCAT 600  
TACTGTGCTC ATGCTCGGA CTTCTGGAG AATGCTTTTG CTCTCTTCT GAGAGAGCAG 660  
TATGATTTGG CTACAGAGAG AATGGGCGAG CTTCCTGACT TAGACCTGTA AGTGGATGTT 720  
CTGAAGGCA ACACCAATGA GGTCTGTGG GCGGTGTGG CCGGTTCAC CAGTAATCTC 780  
TCTCTGGGCT GAACATCTT CTTTCTCTC AAGTAACCA GTAGTTTTC TTCTGTGAC 840  
TTCTGTGTTT CTGTATATTT TACTTCCA CACTATATTT GGCCTCTGTT TTACAAAATG 900  
GTGGTGGCT TTTTCTTTT TGTAGTGTG CAGGATCTG CTGTAGAGAG AGGCTTCTT 960  
CTTCTGTTT TTAATAAAG TTTTACTGCC ATATGGCAT TCCATTCCTT GTTGCATCC 1020  
TCACTGTAC CTTTGTGGG TTTCTGCTCT ACTTGAATTT TCAAAATGAC TCCAGGCTCC 1080  
TCATACGAC AGCTTTTGA TGACCTCAG TTGAGTTTCT CCAATATGTC ATGTACTCT 1140  
AGCATTTGCT CTACAGTTCA GACAGATG ACAAAAAGGC CTTCAACTCA CCAAGGTAA 1200  
ATATCTGAT CTATAGGAC ATTTTITACA TAGACTTCAG TTGAGTGTG TACTTAGCAA 1260  
AATATTTTT AATGTGAAC AGCAGATTA ATACTTAATA TAAATGTCC CTTGGATTTT 1320  
GCTTCCCATG TAAATCTATT GTATTATAC ACTGTATATA ATTITACTA TAAAGTCCA 1380  
ATGTTTTCAC AGAGCCAGTT TGGATGGGC TGCATTCAT TTATGCTGTA TATAGTTGA 1440  
ATATATATA AATTACCCCT TCTCTGGCC ACCCTGCTC CCATCTTAOT ATTTTGCAG 1500

ATCTAATCAG TTGTACACCT GGTGCCCCCT GCTTGTCTCA ATCAATGTTA TTGTATGCA 1560  
AAATGACCT CTGTGCTG AGAGAGAG AAAAGATGTG TGTGTATGTG GTTCTGGAAT 1620  
TTTTGACCT GTGCATTTA TGTACTCT TGTCTATGCA TCCCCCTTTT AGATTTTTT 1680  
TAAATTTAT CTACTGTT TTATATTTT TATGTGAG AGGCTTTGCA CAGTACCA 1740  
TCTGTAGTT CTTTTCTGT CAGAGATA ATTAATATCT GCTGTGAAT GTCAATTAAC 1800  
TACTGACA TTCTTGGA AAAAATCA ATGTCAATCC TACAGATGT TGCATTTAA 1860  
TTGTATGCA GTATGATTA CAGCCAGG GATTAAGAT TTGTATCAG AAAGCTAT 1920  
GTTTATATTT TTATATACA ATTAAGATA TTACATGT CAGACTTAA ACCTTTCTT 1980  
TTTAAAGTT ATTTTACTA TTCTTTATC ACTTTATGT ATCAACCA TTGATTTCA 2040  
AATGTAAAT GTATATGTG AACAAATTA ATGTCAAAT TTTTATATAC CAGATGTAT 2100  
GTAAATAGT GAGCTTCA GGTTTAGAG ATTTTTTGT TGTATTAA CATTCAATCC 2160  
AAAAATCA GATGTGTAT AACTGTAGG TTGAATTTA AGGATTTCC TAAATGTAT 2220  
ACTATCTTT TACTGAGT AATTAATTA CAGATCTT GAAATGCTT AAAAATTA 2280  
AAAAATTA AAA 2293

(2) INFORMATION FOR SEQ ID NO: 186:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1212 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

GAGAGAGAG GAGCCGAGAG ACCGTACCT GGAAGATGTG GTTTCACATC GTGCCCCCTC 60  
CCCTGAGTTT GAGACCTTTA GAGACTATTT GGCCTGCCCC AACCCGAGCTG GAGACATGAG 120  
GCGGTGCTGT CTGTGCTGAG GCGGCTGCTT GGCCTTCAAG CTGTGCTGAG 180  
GTATAGATGT GATTCACCA CCGTTCCAGG CCGAGATGT CCGCAAGAT GGCAGCATG 240  
GAGCAGAGG GAGAGCATGT CTACAGGAT GTTCTGATGT GACTTGAATA TGAAGATAT 300  
GAGATATGAG AAGACACAGA TTATTAAGAT GGCCTGTGTG ATATCTGAT CTGATTTCAA 360  
CATTTTGTCT GAGATCTTA ACTGATTAAT AAAACAACA GATGATGTCT TGAACAGAT 420  
GTCAATTTG GTTAAGAGAG ATCAAGGAAA GTCTGAGCTT ACCAAGCAG TGAAGAGAG 480  
TACATTTACA TTGAGAGAG CAGAGATACA ATTTCTGTCT TTGTATCAG AGCAGATCTC 540

TGCAGGCTCT TGTCACTTGT CAGAGATTTT AGTTCAATGA GATTAAGAT TTCTTACA 600  
ATACATCTCC GAGTTTATGA AACATTTCA TTATGAGATA ATTAATGTGA GCACTGTAAA 660  
AGACTGTG AGACCTGTGT ATTCAGAGA ATTAATATTT GACCAAGGA AGGCTGTCTC 720  
TGTATGAGCA CTGTATGCA TTATGTAAAG CATCAAGAG CTTCAGTTT ACCGAATTA 780  
CATCTTCAAG AAAAATATATG ATGAAGAGA GAGGAATTT ATGAAGATG GCGAATATGA 840  
GAGACCTGT AGTTGATGT CAGATCATGT CTGCACTAC ATGTATATCT GAGGCAACT 900  
TGTGTGTCT TTTTCTCTC ACGTATGTG CTGTGAGAG GACTGTGAT TTACTTCAAT 960  
CTCGAGATGT ATTTCACTAG CAGACAGAC ACGAATATCT ATTTTCTCC TAAATGTCTG 1020  
TTTCAATAT GACAGAGAG CTCTTTTGA AGTACAGAT CATGTCCATC CTTGTGTACA 1080  
TATATCATTT TCTTTTAAA CCAATCTTT TGTATATA AATAAATGAG TAAATTAAGC 1140  
TATGTATAT GAAATGAAA AAAAATTA AAAAATTA AAAAATTA AAAAATTA 1200  
AAAAATTA AT 1212

(2) INFORMATION FOR SEQ ID NO: 187:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1605 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

GTGTGAGAA GTTCTTTTGT TCCAAACATC GGGCTTTCT CTTTTGTGT TCCGCGCAT 60  
GCACTGTCT CTGACCTGT GAGGTCTTAC TTCCGAGAG CCAATCTTG CCACTTCCAG 120  
GCGCGGAGCT AGCGCGGTTT TCAAGCAGAG GAGCCCTCAA GGAACATGAG AACTTCAAGC 180  
GCGCGGAGAG GCGCGGCTG AAAATGAGCT GGCCTGGAAT GCGGAGGTTT GGAAGAAAC 240  
ATTCAGAGAG GAGCTTCAAG TGTGATTTAC ATGAGAGACA TGAATGATAC CTGAGGCTCT 300  
AGCTTGAAG ATTAATGATGA GCTGATCAG CAGCTTCCAG AGGAAGAT AGACCTGAT 360  
GAGATCATGT CAGTGTCTG TGAAGAGAG GATTAAGAT TCTTGGCAT GAGAGCTTT 420  
AAGAGAGAG TGAACGAGCA GTTGGCAT CAGATGTGAG AGGCTGAGAA AACAGAACCC 480  
TGCAGGCTCT TCACTTTGA GCGCAACAT GACATCTCA GAGCTTACTT TGAATGTAG 540  
CGTCTTCAAG TGCAGAGAG GGTCTTGAAG TGCATGATCC CTATCAAGAT GTTCAACTTC 600  
CCCAAGAAA TTGAGATGA ACTGATGA CCTGTCAAG TGGTCTTCA TCTGTGTCT 660  
ATCTTACTCT ATGATATGA GAGCTTCAAG ACTATATCT GGAAGGAGC CTAATGTGCT 720



5 ACAGCCATTG GCACTGCTTT CGGCTACTGG CTGGAGTCT CATCTCTTCAT TTACTTCTTT 780  
GCCTACTGTT GCAGGCCCA GATCAGCATG CTGCAAGTGT TGGCACTGCT GGGCTATGCG 840  
CTCTTTGGGC ATTGCATTGT CTTGTTCATC ACTTATPATA TCCAGCTTCA CCGCTCTTTC 900  
TTACTCTTCT GGGCTTTGGT GGGTGGACTG TCCACACTGC GCATGGTAGC AGTGTGGGTG 960  
10 TCTGGAGCG TGGGCCCCAC ACAGCGGTG CTGCTCTGTG GCAGCCCTGGC TCCCTTACAC 1020  
ATGCTCTTCC TGGCTCTATCT GCATTTTGCC TACCACAAAG TGGTAGAGGG GATCTTGGAC 1080  
15 AACTGGAGG GCGCCACATC CCGGCCCATC CAGAGGGTCC CCAGAGCAT CCGTGGCATG 1140  
CTCCCTGCTG CTGGCTTCC CAGCAGCTTC CTACAGGCCA CAGCCAAAG TGTTCGGGTG 1200  
ACCTGCACT CACTGTACC CCACCTGAAA TTCTTGGCCA CTCCTCTTTC CCGCACTGCG 1260  
20 AGAGAGAGG AAGACTATTA AAGCAGATC CTGATGACAT GTTGTGTAGA TGGGGTTTGC 1320  
AGCTGCCACT GAGCTGTAGC TGGTATAGTA CTTCTTTGAT GCTGTGCGC ACTTCTGAAA 1380  
GGCACAAGC CAGAGACTCC TGGCCAGAC TCCAGAGCTC TCCAGCCAT CAGAAAATG 1440  
25 GGTGAGTCC TTTCAGAAC CCTCCCAACC TACCCCTTCC TTCTCTTTTA TCTCTGCCAC 1500  
ATTGTCTGC TAATATAGA CTTGTATATT AATATGTGA TTGAAGTCTG GAAAAAATA 1560  
30 AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAAA AAAAAAAAAA TCGAG 1605

35 (2) INFORMATION FOR SEQ ID NO: 188:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1516 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

45 ATTGCGCATG AGGGGTGAC GTGTGGCTG GCGCGGGAA ATGGGGGCTT CAGAGAGAG 60  
CGGACTTCA GCGCGCGAG GCAGCAGCA GGAAGCATTT ATGACTTCT ACAGTAGGT 120  
GAACAAATA CAGAGAGAG ACTGGTTCTT AACTTCGAAA ATCAGATTG AAGACTGAC 180  
CGTCTGCTGT TCTCTTACT TCAATTGAA CCAATTGAG GTTCTTACA TAGATCTCA 240  
AGTTACAGAT GAAGAAATA AAGAGAGTT TGGCAGTTA TCCATCTGG TCACTCTGA 300  
55 CAAAAATCA GATGATCTG ACAGAGACA AAGGCTTTT GAAGCTGTG ACNAGCTTA 360  
CAAGTTCTTA CTGATCAGG AGCAAGAA GAGGGCCCTG GATGTATATC AGCCAGAAA 420  
AGATAGCTG GACACACTG TGAAGGCG AAAAAAAAAA TTAAGAGAG AAGGAAAA 480

TACATTTGTA CAGAGAGATG ATCTGAGCT GTTCAAACA GCTGTATATA AACAGACAT 540  
GAACCTCTTT GCAGAGCTGG AATATTAAG GAAGAGAGA GAAGCCAAAG AGATCATGA 600  
5 AGGAAACCA CAAAGGAGAG AAGAGATTGA AGCTCAGAA AAGGCCAAAC GGGAAAGGA 660  
GTGGCAGAA AACTTTGAGG AAGGTGAGA TGGTGTGTG GACAGCTGGC GAACTTCCA 720  
AGCCATPAG AAGGGAGA AAGAGAGAA AATCGGACC TTCTTGAGAC CACTGAAAT 780  
10 AAAAAAGAG CAACTGAGT GACCGCCAA GGTCAAGGC ACAGAACCTT TCCCTCTCTA 840  
TCTGCTTCC TGGTTGGAG GACTCATCTT TTCTGCCAC TTCCAGCCCA ACATAGATTA 900  
15 GTATTTGCTT TTATGTCAT TTGTTTTCA ATAGATTTA ATATCGATCA GAGTAATCTT 960  
TTTGTACTT GAATGAGGG GCTTGGTTTA AAAAAAGAC TTCTCTCTTC CTTGCCCTTA 1020  
GACACACAG TATTAGAGG TCCACCATT GGTGCTGCTT TCTCTTCCA CAGCTGTAA 1080  
20 CTCAGTCTTT TGTACTTAC TGAATGTGA TGGTTAGAA CTTCGTGAT AGTTTGTGA 1140  
AATATCCAA TTAAACATAC TGTATTAAC AGTGTGCTG TCACTTACA GACAAGCTG 1200  
25 GAAGGGCAC CTTAGAGGC CCTTGGCTT CAGTTGCTG CTTCGGGTG TCTCTGCTTC 1260  
GAAGGCCAG ATAGACAGG GAACACTTGT GACACACAG AGCAGCATCT GATGCTCTT 1320  
GGTGTGGC ATGTGCCCC TGTACTTGA CCAATCAGT TGGCATGAG CCCAGGCCAC 1380  
30 CCNAGCTTT CACTTCCA AGAGTAGCC GTCTCCACC CAGTACCATG TCTTAGCCTG 1440  
TCTCATTTG TTAGTGTAA TATCTTTAT GTATATATA TTTTATACC CAAAAAATA 1500  
35 AAAAAAAAA ACTCGA 1516

40 (2) INFORMATION FOR SEQ ID NO: 189:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 681 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

50 GCTCCCATGT TCGTGCTGT CCGTACATCA CCGTGTCCC TCCAGAGGG GCGTACAGGC 60  
CACTCTCTC CTGTAGGCT CTGACTCCC TCCACTTTG GCGCTCAGC TTATCTGGG 120  
CAGGGGACA TTGCAGCATC CTCCCTGCT CNGGACTCAA GGTGCTGAGG TATAAGGCTT 180  
55 GGGCCCCAGA TCCCTGRTKA CACTCTCTG GAGAGACTC TCANAAGTGA CTGTATNTT 240  
GAGTTACCA GCATTAATCT CCGACTCG AAGCAGGTCC AAACCCAGG ATCCAGGCT 300  
60 CCGTGGGCTC TGTGGACTG TCTTCCGAG ATCTCTCTG TTGCACATG GGAACCTAA 360

GAAGAAAAAG AAGAGGAGCCT GCTTGGCCAG CAGTGCAGAG GATTCAGTGC CCACTGCCCC 420  
TCTGTCTGCC TGGCTGGAAAT GTGGGCCCCCT GCTGCCCGTC AGCTTGTGCT GTCTGTGACG 480  
TATGTTTACA TCCCGAGAGG GTTTCATGCT CATTGCCAGC CAGTGCAGAG TGTGTGTCAG 540  
CAAGTTCAGT TGGGTGTCTG ACATGTGTCA CCACTGCCCC CTTTGCACCC GGGGGGGGTCA 600  
TGGTGTCTCT GTGAGTGTCT GTCTGGAAAT TTTTCTTTTCT GAAAGAACCTT TTAAGATTAA 660  
AAAAAAAAAAAAAAAACTCG A 681

15 (2) INFORMATION FOR SEQ ID NO: 190:

20 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1014 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

GGCTGAGCC ACGCATGGA TAAATTTCTG GAACTGTCA ATTGAGTGT TGTACAGCA 60  
GTTAGTTTAC AAACCATTC ACCAGACCA TCAAAACGAA GTCTGCCACT TTTTGTGATC 120  
GCTGAGTAAT TGTCTGGCCA AGGTGACCTT TTAAAGCTCT TAAAGACCCA GAAAGAGGCT 180  
ATTGCTTAAAG CCAAGATGCA CAGATTTGGT TATATGTCAA AGGCCATTCG CTCTATGATG 240  
TATGGGTGAC ACTTGTGAC AACCTGAAAG TTCAAGAAAG GCAAGACGAA GTTGTGACCA 300  
CTGTAGCAAT AGCATATGTT GCAAGAAACAT GTTCAACCTT TACAGTACTC CTTCCTTTAA 360  
TGAATGAATA CAGATTTCTT GAACTGATG TTCAAAATGAG AGTGTAAAA TGGCTTTGCT 420  
TCTTGTTTGA ATATATTTGT GAAATGGGAA AAGACTACAT TTATGCCCTA ACACCTTTAC 480  
TTGAGAGTGC TTTAATGATAT AAGAACTTTC TAAACAGACA GAGGCTATAT GCAATGTGAC 540  
AAGCAATGTC ACTTGAGGTT TATGGAATTC GTTGTGAAAG TTGGGTGAAAT CAGTGTGTCA 600  
ACTATGTAT GCGCAATGTA TTGAGACAT CTCTCATGAT AATTCAGGCA GTTATGGGAG 660  
CCCTAGAGAG CCGTAGAGTT GCTATGTGAC CAGTGGAAAT GTTGCAGAAAT TGTTTACAG 720  
GTCTGTCTCA CCGAGCCCGG AAGTGCAGAG ATGTATATTC GAAATTTTAC AATCTCATCT 780  
ACATATGTTT CCAAGAGCTT CTCAATAGAC ATTACCGAG AATTCAGCA CGATGATAG 840  
TAAATGCAGS TACTTGCAC CTTAAGTTTC GTTGAATTC GTGATGTAAA CTTTAAACA 900  
TTGCAGATCA GTTGAAGACT GTTGCATAGS GAAAGACTA GAAATTCAT AGGC 960  
1014

5 (2) INFORMATION FOR SEQ ID NO: 191:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2779 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

TGGCAGCAGS GTGTGTCCAG ATGTGTGATC TGTGTGTGCT AGCTTGTCTT GACAGGGAG 60  
AGTTAAAGTC CCGTGTCTCA CCGTGCAGGC TGGCCAGATG GCTATAGGCT GACCGAGAGA 120  
CGAAGACTTG CTTCCTGAG AGTATGTCTC AAGACTTGGAG AAGGAGGTTT CCGCCCTCTT 180  
CTCTGTGAG CCGCGGGCAGC CCGTCCCGGC TTTCATTTCT CCGCGAGCCC CTCTATCTGA 240  
GAACTTCGAG GATCCAGACA GCGCCAGACG CCGTGCCTCA GCTTGCAGGAG CTTCAGTCTAG 300  
GCCAGACGCG ACGCGCATNG GAGAGAGAC AAGAACCTTC AGATCTCAT CTGCAGAGAG 360  
GTCTGTGATG GACGAGAGAG CCGTCTCTCT GTAGATAGAC CTGACCTCTT AGCTTGTGAG 420  
TTTTCAGGTT GAGAGCATTA GATGAGAGCC AAGAGATNG CTCTGAGGAG GACAGAGGAA 480  
AGCTGAAATT TGGAGAGCAG CTCTCTCCA TGGATGTACA GTTTCAGAGCC GAGAGACGGA 540  
AATTCGCCCC TCGATATAGA GTTCAGTCA ACTTACGAAA GCGAAGAGCT GCGATGTGAC 600  
CGATTCGAAA CCGATTTTAC CGAATTCAGC TCTTTCATNG GGTTCGCGCG GGTTCGCGCG 660  
AGATCTGAG TGGATCTTCA GATATCTGA GCAAGACAG CAGATATCTC ACCGACTGAG 720  
AATACAGACA GGGCTGCACA GATTAAGAT GCTGTATGAA GGTGTGTCTG AACCTTAAAG 780  
ACGGATGCAA TGCCTGCAAT CTGCAGTGC TGCAGATGCA CCGGAGCTT GCGAATCTTC 840  
AGCCCTGCTT AATATCCAG TGCAGATNG ACTATTTACAG AAGCCAGACG GCTTGTGACA 900  
TCCCATTTGA GAAAGAGAGI CTGCAGTNG TAAAGCTCTT GGTGAGAAAT GGGGGCAATG 960  
TGCATGCGCG GATTCGCGGC GATTCTTCA GAAAGGCCAA GGAATTTGCT TTTATTTGAG 1020  
TGAAGTACCC CTCTTTTTCG CCGCTTGCAC CAGAGATNGS GATGTGTGAA GCTACTCTCT 1080  
GAGAAACCAA CAGAGCGCG CAGAGCTTCA GGGCATTAAT CCGAAGGGCA CAGATGTCTG 1140  
CATGCGTAG TGAATATCTC GAGACATCA GCTGAGACA TTGCATGCTT GAGCAGCATG 1200  
TATGATGAGC TCTTCCAGAC TGGGGCCCCC CTTCCTCTTA CCGTGCAGCT TGAAGACATC 1260  
CGAAGCTGCG AGATGTGAC GCGTGTGAGS CTGGCCCCCA AAGAGGGCAA GATTCAGAAAT 1320  
TTCAAGACA TCTTCAGAGS GAGATTTTCA GATCTGAGCC ACTTTCGCGS AAGGTGACG 1380  
GAGTGTGCT ATGGAGCTGT CCGGGTNGS CTGTATGACG TGGCTTCTCT GACAGAGCTT 1440

5 GAGGAGAACT CAGTGTCTGA GATCAATTCG TTTCATTTCA AGAGCCGCCA CCGACACCGA 1500  
ATGTCTGTTT TCGAGCCCTT GAGAAACTG CTCGAGCGA AATGGAACTT GCTCAATCCC 1560  
AAGTCTTCTT TAACTTCTT GTGTAACTG ATCTACATGT TCTATCTTAC CCGTGTTCG 1620  
TACCAACAG CTCACCTGAA GAGCAGGCC GCGCTCAC TGAAGCGGA GGTTCGAAC 1680  
TCCATCTGTC TACCGGCCA CATCTTATC CTCGTAGGGG GAATCTACT CCGTGTGGC 1740  
CAGGTGTGT ACTTCTGGG GCGCAAGTG TTCACTCTGA TCTGTATTAT AGACAGCTAC 1800  
TTTCAAAATC TCTTCTGTT CAGGCCCTG CTCACAGTGG TGTCCCAAGT GCTGTATT 1860  
CTGGSCATCG AGTGTACTT GCGCTTCTT GTTCTTGGC TGTGTCTGG CTGCTGAC 1920  
CTGCTTTACT ATACAGCTGG CTTCAGCAG ACAGGATCT ACAGTGTAT GATCCGAG 1980  
CGCTGTGAG CCGTAGCCAG GANNITGGG CCGCGAGCT CTCACAGGCC CCAATGCCAC 2040  
AGACTCAGTG CAGGCCATCG AGGACAGGA KGAAGAGGCC AACGGGGCC AGTACAGGG 2100  
TATCTCTGAA GCTCTCTTGG AGCTCTTCAA ATTCAACATC GGCATCGGGG AGCTGGCTT 2160  
CGAGGACAG CTCGACTTCC GGGCATGCT GCTGTCTGTG CTGCTGGCT AGTGTCTGCT 2220  
CAGCTACATC CTCGCTCTCA ACATCTCAT GCGCTCATG AGCGAGACG TCAACAGTGT 2280  
CGCCACTGAC AGCTGAGCA TCTCGAGCT GCGAAAGCC ATCTCTTCT CTCGAGATGA 2340  
GAATGCTTAT TGTGTGTCCA GGAAGAGCA GCGGCGAGT GTGATCTCTA CCGTGTGCC 2400  
TAGCCCAAT GCGAGCCGSG ATGAGCCCTG GTGCTTCAAG GTGAGGAGG TGAATGGCC 2460  
TTCATGGAG CAGAGCTGTC CTAGCTGTG TGAAGACCG TCAGGGGCGAG GTTGTCTCTG 2520  
AACTCTCGAG AACCTGTGTC TGGTTCGCC TCCCAAGGAG GATGAGATG GTGCTCTCTA 2580  
GGAATACTAT GTGCGCGTCC AGCTCTCTCA GTCCAACTGA TGGCCCAAGT CCGAGCAGAG 2640  
GCCAGAGGAC AGAGCAGAGG ATCTTTGAA CCACATCTGC TGGCTCTTGG GTCCCAATGA 2700  
ATTCTGTGG CAATATATA TTTTCACTTA CTCAAAAAA AAAAAAAA AAAAAAAA 2760  
AAAAAAA AAAAAAGC 2779

(2) INFORMATION FOR SEQ ID NO: 192:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1923 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

ACCGCTCCG CTCGCTCG CGCGCCCGT CAACATGATC CGCTCGGCG 60  
TGGCTGCGA GCGCTGCGC TCGATCTCTC CCGTCTCTCT ACTCAGCGC ATCGCCTTGG 120  
5 AACTCATGCG GCTGCGCGCG CCGCGCTGCT TCGAGTCTAG CGACCAAGCG CAGAGTCTCT 180  
CGCTGTGTG GAAATGCTCC CAGAGGGCG CCGCAGCGG GTCTTAGGAG GAGGCGTGT 240  
AGAGCTTCAT GAGTACGGG TGGGTAGAG CAGCGGCTGC CATCTCTTTC TGTGCTTCA 300  
10 TCACTCTGCT GATCTGTTC ATCTCTCTCT TCTTGGCCCT CTGTGAGCC CAGATCTTTC 360  
TCTTCTGAG AGTCAATTGA GGTCTCTTTC CTTTGGCTGC TGTGTCTCAG ATCATCTGCC 420  
TGTATATTTA CCGCTGAG TACACCGAG CTTTCACTCT TCAATGCCAC CCGTCTGTCA 480  
CTTACATCTA TAACTGGGCC TACGCGTTTC GGTGGCAGC CAGCATATAT CTGATTTGGT 540  
GTGCTTCTT CTTTCTGCTC CTGCCACTT ACGAGATGA CTTTCTGGCG AATCGCAGC 600  
CGAGTACTT CTACACACTT GCTTAACTTG GGAATGAATG TGGAGAGAAA TCGCTCTGTC 660  
720 TGAATGAG TCACAGAGAA GAACTGTCTT CTCAGCGCA CTTTGAACCC ATTCTTTGGC 780  
AGTCTCTATA TATTAAGCT AGTCAGAAAT GCTAAATATA TTTGGGAGAA AATATTTTTT 840  
AAGTATGTT ATATTTCTAT GTTATCTT TATATGTTT TGTGAAGTTC TGTCTTTTCA 900  
CTAATTAAT ATACTATGCG AATATTTCTT TATATCTATC CATTAACATTT ATACTACTT 960  
TGTAGAGAA TATCGAGCTG AACTTTAACA CTTTATAGG TAAATATGAG GTTTCACAGA 1020  
TTTATATAT TCACTCAAGTT CTGTATATT CCANATAGAA TGGACTGCTT CTGTATAGG 1080  
CTAGGAGAA GAGGAGATA AGTTTAAAG TTGTATATCA CCANACTTC TAAAGAAAT 1140  
CGAAAAAA AGTTTATTTT CAGCCTTCT AACTTTTAA GGAAGCGAA ATCATTTCTT 1200  
AATGCAAT CATTTCTGAG AATTTCTCAT TAATATCTG AATCATTCAT TTCAGCTTAG 1260  
CGTCAATGTT GACTCGATAT GTCATCTAG AAGTACTAT TTCAATGGTCC AAACCTGTG 1320  
CCATAGTGG TAGGCTTTC CTTTAACTGT GAATATTTA GATGAATTT TCTCTTTTAA 1380  
AGTCTTTAT AGGTTAGGG TGTGGGAAA TCTATATTA ATAAATCTGT AGTGTTTTGT 1440  
GTTATATGT TCGAACCCAG AGTAGACTG ATTGAAGAT GGACTGGGTC TAATTTATCA 1500  
TGACTGATAG ATCTGTTAA GTTGTGTAGT AAAGCATTAG GAGGTCATTT CTGTGACAA 1560  
AAGTGGACT AATACAGCT CAGGAGATA AATGACTTGC TTTTCTAAT CTCAGGTTTA 1620  
TCTGGCTCT ATCAATAGA CAGCCTTCTG ATAGTTTCA ACTGTAGCA GAAGCTTCA 1680  
55 TATAGTAAA ATCTGTGCTT TTCTGGTAA ACAGATTTTA AATGTCTGAT ATAAACATG 1740  
CCACAGGA ATTCGGGAT TTGATTTCT CTGATPAGA TATATGAT GCATCGATA 1800  
GCTCATATG ATTTTATACC ATTTGACTT ACATATGAA AACCAATCA TTTTAAATAT 60

CGAGTATATG TTTTCTAATG TGTGAGAAAA GCTATATGTA GTTTTCAATG TGAAGTTTC 1860  
CGAATTAACC AGGTATTTCTA AAAAAAAAAA AAAAATACTN GAGGGGGGGC CCGGTACCCA 1920  
ATT 1923

10 (2) INFORMATION FOR SEQ ID NO: 193:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2146 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

AGCGTCAGAG GAGACATCTC AAAATTACAC AGCTTTTAC AGGTGACAAA ATTGGGCTTC 60  
AGACCGAAT CTGGATTCAA GTTCACTGATG GTTATGTTGC GGCATTCTTT CCGGCATCTG 120  
GAGCTTACCA TCTCTCTCTC CGATGACAAA TGAAGAGAAC GGGGGCCCCG CAGCTGATATG 180  
GCTGAGAGAG ATTAAAGTTCTT CTCTGGGGGCT GGGCAAGGTAT AAGAGCATGT GGTCTGTGGA 240  
CGGCAATCCT GGTCTCTCAC CTGGCAGTCT GCTTGCCTTG GTAAACGGCT GTCTCTGTCT 300  
CGCTTATGTC CCTTGGCTAG CATGACCCGC CTGATGCGAT SCTGACAGAC CTCTGTCTTC 360  
AGCGTCATTC TGTGATATGC ACCGCAAGCC GCTTCCACAC CAGCGAGAGCC ACCGCAAGCC 420  
GCTTCCACAC CAGCGAGAGCC ACCGCAAGCC GCTTCCACAC CAGCGAGAGAG GCGCACTTGG 480  
AGATCAAGCC CAGCTGAGGT GCTGCTGAGAA ACAGAGCCGG GCGCAAGTCC ATGAAAGTCT 540  
CGTGTATAGC GGCCTGACCA GCTGCGGGGCC CCGGACATCTG ATCTCTTAG TGGCCCCCTC 600  
CTCCCCGAGC CATTCTTCCC GCGTGCCTGC CATATCTGAC GTAACTCGAT ATTATATCAA 660  
AGCTTATTTT GTAAAGATGA GCTCTGATGG AAGAAATATG GATTATTTAC GTGGTATCCC 720  
TCTGCAAGAG CCGGATGAGG GTGAGACAAA GGAAGAGAGC AAGCATCTCC GATATGATTC 780  
CTTCTCAATT AACGATGAGC CCGATTGACAC TCTCTCTCCC TCCCTGAGAG AACACCAACT 840  
GCGAAAGACA AAGACGCTAC AGCACAACAT TCAAGAAAGC AAGCTAGAGC CCGCTTAGAC 900  
ATCTCTGTTC AAGCGGTGAC CTGTGTCAGAA GCGCAAGCCGC CAGTCTCCCG TTTCTCTTTT 960  
AAGTGAAGAG AAGCTGATCC AGTTTCCGGA AACAAATCCC TTTTCTCATT TGGGAGAGAG 1020  
GATTATATG AGATGAGAGC ACCTCTTTTA AACAGCAA AACAGAGAG GAAAAAGTGG 1080  
GAAATCAATG CAGAGCTAGA GGCATTGGA ACAAGCAATC TACGTATGTA ACTTGAAGAA 1140  
ACCGATTTT AAAGTGGTGG CATCTAGAAA GCTTTGATG CAGAGCAA CAGCTTAT 1200  
TTTCTAGCA TCTCTTAAAT GTGAGAGAA AGCAGGAC AAAATCTCT GCTTTAGAG 1260

ACAAAAATAT TTGACAGAAC GTTGGGATAC ATGGTTTGG AAGCTTTGG TTCTCTTTC 1320  
TCTCTCTCT CAGAGAGCCC AACCTCCAC CCGTATGACA TGAACAGATG ATTATCTTGG 1380  
TTGAGAGAGA AATATCATTA GATTGTGTTT GATTGCTTGA GATGAGAGG CAGATCTCAA 1440  
CAGCTGCGCT GCTGTGATGG ATGTCTCTTG CAGGGGGCGG AATGAGAGCA CTGGGGTGGG 1500  
GCGGAAATGG GGGTATCTGG ATGTAAAGGA TTCTCTTGG TTGTGTGAG ATCCAGTCCA 1560  
GTGTGATTT CTGTGATGCC CAGCTTGATT CAGAAATTT TGTGTGATG GCTTAATTC 1620  
AATTTCATG CTTCAGACAC TGGGCTGAAA CTTGAATCTA GTACTGAC CTGTCTGAC 1680  
CGATCAAGTT CTGGATCTCT CAGAACTTT TCTCTTGTG GGGGTGGGGG TGGGACTCA 1740  
CTTGGGAGAC GATGCTGAG AAAATGTAG GATTCTGAAA TGAATATCC ATGCAACTTT 1800  
CCTTCCCTCT GCTGCTCTCT CTTTCTCTGC TCCCTAATCT TTGCGCAAT GGGGCAAGCAC 1860  
CAGTAAAGTT TGTGGGGGGC CAGTCCGGCT GCGAGTTCC TGTACTACTG CATTGTACTT 1920  
TTCAATTTGG CTGACCTGAG ATTTCATAT AAGAAATTTG GTGAGATGA ATTGAATATT 1980  
GTAAATGAC CAGTGGAGCC CAGAAATTC TGGAGCCCG CAGTGGAG GAGGAATGAG 2040  
TCAAGCTTTC CAGTGGGCTT GAGAGGCAAT GACTGTGAG CTGGGCCCCA TCACTGTGGA 2100  
GAGCTTCCCT GAGCTTGAAT AAAAAATGCG GGAATGCGAG CAGAGAGAC TGAATGAGGA 2160  
TGGAAACTTA TTGTGACAAA GTCTTTCCAG AAGAGTTTCT TAAATGATTA TTGTATTTTA 2220  
TTTGCAGACC AATTAATTTG TAACTTTGCA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2280  
AAAAAAAAA AAAAAAAAAA CAGGGGGGGC CCGTACCCA TTGCGGTAT ATGATCTTAA 2340  
AAGATC 2346

(2) INFORMATION FOR SEQ ID NO: 194:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3054 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

TATCTGACAC ACCCTTAAAT CTACATATGA TGAAGAGAC TGAATATCC TAAACCCCTA 60  
AGCTTAAGAT GCGCTGTGAT AACAGCAACT GAGATTAAGC AAGGCTGGGC TGTGTCTTCC 120  
TGTGTATAG CTGACATCTT CCGCCAAATC TTCTCTGGA GCTTCTGAC CAGCTGTCTA 180  
AAGCAATAG GTTGAAGACC TCTTCTATCT ATCTCTTAAA AGCAATATGT TGAATCTTCA 240

425

426

TTCAACAGCT GATGCCCTAT AACCCCTGCC TCGATTCTTT CATTATAGGC TATTAGAGCT 300  
AGCAGAGTCT TTACATTAAT CAGATGGGTT TCAATGGCTT CCTACCCCTCT CTAATGCCCC 360  
5 CTCCATTATT TTGACTTAAG CATTACACAG TGGCACTAGC ATTATACAA GAGTATGAGA 420  
AATACAGTGC TTATATGGTC TACATTAAT GCTTTCAGTA TCAAGGCTGC CTGAGGAAG 480  
GATGGAGCC TGAAGGCTTC CTTATGTCTT CCACACAG AGCTGCTTCA TGAAGGTCAAT 540  
10 CTTTTCCTCC TATCTCTGTC TTCCCTCCCC GCTCTCTAAT GGTATGCTGG TACCAAGGCT 600  
GGTCTCTGG CTAGTATAGT GGGACCAAGT TCAATACCTC CTTATACGTT CTAGCATAGT 660  
15 AAACCTAGCT ACCAGTGTGA GTGGAGAG CTGGGTTTTC CTAGTATACC CACTGCATCC 720  
TACTCTCAGC TGTCTAAGCC GCTGCTTCCA GGTATGGGAC CTGCTAAGTG TGGAAATACC 780  
TGTATAGGA GAGGGAATA CAGAGGGCC CTCCTGTGTT CCTGGCTTCA GGCAGCTGCC 840  
20 CACAGGCCAT AACCAATAA AACAGATA CTGAGTCACT TTTTATCTG GGTCTCTTTC 900  
ATTCTCAGT CACTGTGTGC TGTCTTGGCT GACTGGGAAC ACCCAATAC TACAGAGTCT 960  
25 GACAGGAGA CTGGAGAGTG TCCATCTCTA GCTGGAACT TACTGTGTAA ATTAACCTTTC 1020  
AGACTGCTA CCATGAAGTG AAATGGCAC ATTTCGCTTT ATAAATCTTA CCATGTGG 1080  
GAAAACTGCT CTTTTCCTCA GGCCTTCCA GGGCAATAAA CTCAGCCCT TCGATAGCAA 1140  
30 GTCCATCAG CTTATTAAT TTTTAAGAA AACTTGCACT TGTTTTCTT TTTACAGTTA 1200  
CTTCTTCTT GCGCCAAAT TATAACTAT AAGTGTAAA AAAGTCTTA ACACAGCTT 1260  
35 CTGCTTGTGA AAATATGTGA TTATACATCT GTATTTTGA ATTCGTCTCC TGAATAATGA 1320  
CTGTCCCAT CTCCACTCAC TCGATTGGG GCTTTTCCA TTGATCTGCA TGTCTTTTAT 1380  
CATTCAGGC CAGTGGACAG AGGGAGAGG GAGACAGGG GTGGCAACA CTGTGTTC 1440  
40 TTTCTGACT ATCTGTACA AGAAGAGTA ACATGAGGC TCTGTGCTCC ATGCACACT 1500  
CTCCAAACA CTTATCTCC TCGAAGAGTG GCTTTTCCAG GGTCTTACT GGGAGCAT 1560  
45 TAAAGCCCT CCTACCCCT TCTTTTTC TTTCTTACT CTTTGGCTT CAAGATTT 1620  
TGAAGAGA ACAATATCT TTACACTCAT TTTCAATTC TAAATTTCA GGGGATCTG 1680  
50 AAATATAGC CAGTGGCTT AAGCTCTGT TAAAGTTGAG GGGAGAGGAA ATCTTAAGT 1740  
TACAGATTA AAAGAGATC CCTTAACAA AAGAGCAAT AGACTGTCT TCTCTTTG 1800  
CCACTTTCC TGTTCATGAC AGTACTAAC CTGGAGAG TACATTTCA TTACCAAG 1860  
55 AAAGTGGTC ACCTGAGCT TGAAGAGTG AGTACTCAG CCACTCCAT CACCTTACA 1920  
GATCCCAAG AGTCCCAAG ATGTCAGCT CTTTAACCT AGCTTAGTCA ATTAACCTGG 1980  
GCAAGTGGC CAGAGAAAT GAGGAAGAT CCACTGTGA GTCAGAGCC AAGATGAAA 2040  
60

GACAAAGAG GAAAGAGTA TCAAGGCAG AAGGAGATC ATTTAGTTGG GTCTGAAGG 2100  
AAAGTCTTT GCTATCCGAC ATGTACTGCT AGTACTGTGA AGCATTTTAG GTCCAGAAAT 2160  
5 GGAAGAAA ATCAGCTATT GGTATATATA TATCTCTCTT TCCCTGGAGT CAGTTTTTTT 2220  
AAAAAGTAA CTCTTAGTTT TTACTGTCTT AATTTTAAA CAGAGGGGAG CTGAGCCAT 2280  
10 TCCCTTAGG AGTAAGATA AAGGATAGG AAGATTTCA AGCTCTAAT AGATCACAG 2340  
CTTTCGAGG TATTAACCT AAATTAAGA AGTACATTA GAGAGGTGG AATATGCTT 2400  
AGTTCTGAT AGCTACCCAC AGAGCAAGTG ATTTATTAAT TTCAATTTCA AACTACTTTC 2460  
15 TTATATCAC TTGTGCTCC ATTTTTCOA GGACAGGAA TATGTCTCC CTTAATTTTC 2520  
TTGCTTCAA AATTAATATC CAGCTTCCA AGATCATCTT ACAGTAAAT TTGCACAGAC 2580  
ATCTCTCAC CCAAGTGGCT GTCTGGAGCT CACCAAGGT CACCAACAA CTGTGTGTG 2640  
20 AACCAACTG CTTTAACCTT CTGGGGGAG GGGATTAAGT AGCTAGGAG ACCAGAGTG 2700  
AATGGAAAG GTGAGGAGT TCACTATGTT GGCCTGTGAG AGCTTGATTA GAAGCCAGA 2760  
25 CAGTGGAGC AAGAGAGAC TTGCCCCAG AAAAACCTGT GGTGTGTGCT AATTTCTGTC 2820  
CAGAAATAG GTGTGACAGA AGCTTTGTGG GTGCATGGAG GAATTTGGAC CTGTTATGT 2880  
2940 TGTATCTTC GCACTGTGAA TTTGTGTGAT GTAAAGACA ATATCTGTGA AACTTAATGT 3000  
CTGTATAAT AATGAGCTT AACACAGTAA AATATTCAT AAGAAGTCAA AAAAAAATA 3054  
AAAAAACTG AGGGGGGGCC GGGTACCAA TTTTCAAAAT AGAGATGTA TTAC

## (2) INFORMATION FOR SEQ ID NO: 195:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 907 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

GGCAGAGCTC GTGGCCGAAA CTTTCTCTGC TCTGTGCTGC CACTACTGCT CTGGCCGGCG 60  
CCCTGGCTG GGCCTGCACC AGCCTGGAG CGGGCTTCCA CAGACGCCCC CTTCCAGACA 120  
GGTGTCCAC ACCGGCAC CTCTGCGGGA AGTGTCTGCG CGTGGCGGG ACCATATGGA 180  
CGAAGGCTT TGTGCTCAC TACAGCTGG GTGAGCAGGG TGGCAGCAGC CTGTTGATCC 240  
55 TCTGTGCTCC TGTGTGAGCA CGAGCGGGT TTTCTGCTCC GAGTTGGGAC TGTGGAAATG 300  
TGTGGTGTCT GTGTGTGCTT CATTGTGCTG GTTCTCCCTG GGTGGGACCT TGTGGCCAA 360  
GCAATGAAA CTGTGCTTC TGTGAGTGC GTGCTGGCTT TCCGCTTCGG GGGCTTAGCC 420

480 TGTGAGACTG CCGTGTGCTT CCACTGTGGA CAGCCGTGAG GCGAGCATG AGCTGGAC  
540 AATCTTGAGA GAGTCAAGCT TGTGAACTT ATGTGTGAGA GACTGTGGA GAGCTCTGAT  
600 CAGCAGAGCT ACCTTCACTG GGAATGATCC GCTTGACGCA GCTGGCCCCC AGGAGCTTGC  
660 AGCCAGACA CTACAGCTTT CTGGCAGACC TGGAGTGTCT GGGGAACTG CTGTGAGGCA  
720 CTGTGGGAG AGGAGCTGAG TGAATGGTTG GAGCAGATC CTTGCTTTT GCTCTGCTC  
780 ATCTGTCTG CTTTTCCTGCT TGTGACTG GAGCTGACG CAGAGACTT TGTGTGAACT  
840 GAGTGGCTG AGTGTCAAT AAAGCAGAT GTGCTGTGAG CCGAAGAAA AAAAAAAAAA  
900 AAAAAAAAAA AAAAAAAAAA GAGGGGGGGC CCGTACCCA AATGCCGGA TATGATGCTA  
907 AAGCATC

(2) INFORMATION FOR SEQ ID NO: 196:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

60 GCGACAGAGA GCGACAGAGA GTGGGAGAG GAGAAAGACA GCTTATTTGA CTACAGACC  
120 CCGTGTGCTT CAGCAGAGCT CTGGCTGTTG TGGAGAGGCT CTGATGACG AGGAGCCCCG  
180 KAGGAGAGA AGAATGGAG GAAAGCTGAG CAGTGGAGAG CAGTGGAGCT AGCTTGTGGA  
240 TGTGAACTA GAGCTCTGCT AGGCAAGCTT CTTGAACTT GAAATGTGA ACCGAGACC  
300 TTAAAGCAG CCGTCAAGCA TGTAAATGAG TTGAATCTAG TGTAAAGAA TATTTATACC  
360 AACTCAACT ACATGTGAGC TGAATGACG GACTAAGCTT GAAATGGAGG TGTTCAGACC  
420 TTGAGAGAGA TGGCAGAGCG GTGCCCTGGG GAGTGTGACA GCGAGCTTAT CTTCTCTGTT  
480 TGGCAAGCTT GAGCTCCGAC CTGCTGAGCC CCGATGGAGT GAGCCCCGCC GTGTGTGTT  
540 CTGTGATAC GTGTGATCC AGCTAAAGG ACAAAAGACA ACCGTGAGC CAGACTGGA  
600 AGCTGCTGAG AGAAGCTCC GAGTGTGAG AGCTGAGACC CTGGAGATG GATTCCTGAG  
660 TGTGCTTAT TGTGTGAGA TGTGTAGAG GCTGCCCTAT AAGAAATGA GACTGTGAT  
720 GCTATATTA AAGACTTTT AAAAGCAAAA AAAAAAAAAA AAAAACTGA GAGGGGGGGC  
780 GGTACCAAT TGGCCAAATA GTGAGTGTGA TTACATTTCA CTGGAGCTGCT GTTTTACAAA  
840 CCGTGTGAG TGGAAAGCT CTGGCTTTTA CCGACTTAA TGGCTGTGA GAGATCCCC

900 CTTCGGCAG CTGGCTTTTA TACGAAAAA AGCCGCGACC GAAATGCCC TTCCGAGAG  
960 TTGGGAGC CCGTAAAGCG GAAATGCA AATGTAGAG TTAAATTTT TKTATTAAT  
1020 TCGAGTTAA AATTTTGTG TAAATCACT CAATTTTGT AACCAATTA GSCGAATC  
1080 CCGCAATCC CCGTATTTA TTCCAAAAA ATTAACGSA AAGGGTTTG AATTTTGT  
1140 TTCCCATTT TTGAAAGCA AATGCCCCC TTTTAAAAA GTTGAGACC CAGACTTCC  
1200 AAGGGGAAA AAAGTTTTT TTGGGGGAGA AAGGGGGGCC GATCAATTA AGATCCCCC  
1260 CCAACATTT TTTTGGGAG GTCCGAAAG GTCCCTTTA AATTTTTTT CCGAACCTTA  
1290 AAGGAGACC CCGATTTAA AATTTAGTT

(2) INFORMATION FOR SEQ ID NO: 197:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1020 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(\*1) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

60 GGTGTGCTG GATGTGCTG TACGTAGTT TTACAGACA TTATGTAGC AAATGATGA  
120 GAGCTTAGG GAGAAATAT TGAATCAAT TAAAGAGAC CTGATAGAG GAGAAATAT  
180 SCATGAGAG ATTGTATAG ACCTTTACA GTATCATTT TTCACTGCA GCGAATGCT  
240 TTTCAGCA ACTTCAATC ATTAAGGCT GTATGAGAT CTATGAGTA GTACAGAGC  
300 AGCTTAGAG TGTATTTCT TCTTTAGAC ACTGTATTT CTACAGAAA TTTCAGAGC  
360 ATTATAGAG TACAAAAAA TCCAGAGC GTTTAAGA TGTGAGTCC CATATATAG  
420 TATGTAGAC TGAATCTGT AATGTATCA CTTTCTCTT TTATCTCTT AACCAAAAA  
480 TTGTTATTT TTGATGCCA AATGTTTTA ATCTGTGAT ATTTTTAAA AATCTTTTC  
540 TCGCATCAT TGCCTTTTT GTGTGTGAA AAGACTTAC TTGCCTTTG AAGATGAGT  
600 ACTGCTGTC ATCTTACAAA TATGTATAT GTATATTTT ATAAAGAGT TGAATTTGA  
660 ACCAGAAAT GTGATTTGT TTAAAGAAA AAAGCTGCT TGAATTTGT GAAATGCTC  
720 TTGAAATTT TCTTTTACA GGTATAGCC AGCTAGATA CCGTATGAG GTTGAATTT  
780 TTCAATGAG CTACCATCT ATTTAACA CTAGACTTT TATTTATTT CTATTTTGA  
840 AGTTATTTG TCTTACTCA TTTAATTAAT ATACTGTTA TGTGTCTG AATGGAGCT  
900 GAGCTTTTG GATATGATA TTGATTTGA AATATTTTG AATTTTCTT ACTTGAATTT  
960 TTAAATATC AATGAATAT TGTATATGT ACTGAAGTA AGTTGTGTA CAGTGAACG

429

430

TCTCTAATAA TATGATGCTT TCCCTTAATAA GAGGAGGAG ATGTCCCACT TTCACACAG

1020

5

(2) INFORMATION FOR SEQ ID NO: 198:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 524 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

AATTCOCGAA CCTCAGGCTT GTGTGCGATC GGGCAGGCA AGTCTTTTGA CCGGACCTT  
CCCGGCGCAG AAGACTGAA GTTCAATTTGA GAGCTGTGTT TTGGGTTTGA GCGAGGCTC  
TGGGGGCTTT GTCCCGGCC AGGACACAG TTACTTTGCA CCGGGCGGCG CTTGGCTTTAT  
GATGTTCTC AACCCAGGG GCGGCTCTGC CCTCTACTCG TCCAGAGGCC ACTTGCAGG  
CAGAGGCTCT CCCAAGGCT TAGAGGCTGC TCGGATCTAC CTGTTGGGAT GCACTAAGAG  
GACCTTTGTC TGGGACAGG TGCTCAAAC ACCCTGCTGC TGGCTGCGAG GCGAGGCTTC  
TGGAGGCA GGGCAGGAC TCATCAGGAC CTCCTCGGAC CTTCCAGGCG AGCGAGTTGG  
CCCGAGGCA AGCAATTTGG TCTGTTGCC CCAAGGGGAC AGGAGGCTC TTGGGCTCT  
TCCCTTCTG GACAAGGCC CCTGCTTTTG CCTCACTAA ACTG

35

(2) INFORMATION FOR SEQ ID NO: 199:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 332 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

GTGATACAG GAGGCTGAT CATCATCTGT CACCATGCAA TTCTGTGCTA CAGCTTTCT  
GTTGTGCCA CTTCTGGCTC TTGTGATGTT CCCCATATCC CTAGGCTTCT CCCCCTGCTA  
GAGGGCTTC TTGATGATTT AGAATATAG AATGCTGAC ATTCTCATG TCCATATAG  
AAGGAGCCAC AAGACATGTC TTTTAATTA AAGGACAGTG TCCATCTTT TAGCTGCCA  
ATAGACCTT GGTCTCATCC TCTGGAGCT AGGCTTAAA ACAGCTTCTG TTTTCTTAT  
TTGCTCATP GTTTTGCCAA GCTTTATTC GG

60

(2) INFORMATION FOR SEQ ID NO: 200:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 376 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

CCAGGGAGC CCAGGCTGT TCTGATGTC ACATCAGTGC TTCCCTGAA C TCCCTCCCC  
ACCCCTGGC ATATCCGAG GAACCTTATG TTTTCTAGAA GCTAAGCAGC TCTCGGACT  
CAGGACTGG TCCAGTAGG CTGATGCGCA GCTCAGTCTT AAGAGGTCTC TGAAGATCTG  
GACTGAGAC CTCTCTACTC CCAGAGCCAG AGCCATCAG CCAGGCTTGC TTGAGGCAC  
CTGCTGTGG ATGCTGAGC TCACCAAG GCTGGCAGC TCTGGGCTC ATTTAAGGA  
TTCTGATG CCGATGGGCC CTGAGGCGAG CCGATTAAAG CATCTGGCTC GTTTTGGAA  
AAAAAAAAA AAAAG

25

(2) INFORMATION FOR SEQ ID NO: 201:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1192 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

CCAGTATAT TTCTATAACA TTTATTTTAG TGAATTTATA ATGTTTCTTT GTATTAAATT  
ATTAGATAT ATCTTTAGAT AATATGTTA CTAAATATGT AGGTAAATTA TATTTTATTC  
AAAAATAT TGTGATCTA ATGCTACCA ATTATGTAC TTGTAGATGT ATCTATCTT  
AACTGAGTC TTCTGTCCT CTAAATGAGT GTGAAGGACT CTTCCTCCT GGGGAGTTT  
TTCTTTTCA GAGGGAGGA GGGCTTTCC AGGTATGTC TCTAGAGTGT TGGCGAAGT  
AATCTGGAC CAGACACAC CAGTCTCTC CTAAATCCAC GTCATTGCC TTCTATCCCA  
GCTATCTTC CAGTGTCTC TGGGTGTTT CAGAGCCAC AGGAATGAA TAAATCTCTG  
ATGAGTTT TATTTTCTT TCACTTGT TTACACTGTA WTTCTGAGT TTATGGGTG  
CTGTGATTA AAAAGGAAA GTTGAAATTA GTAAACTCA GGTCAAGCA AATATATATA  
AATAGATTA AGCTGACTG TAGATATAT CAGGTTATTA AGCTTAGAG TTGCTAAGT  
TGTGTGAAA KTTTCTCTG ATCTTTCTGA TCCGAGACA AAAAGGCG TCATGTTTGT

60

WATGTGATG GAAATGAAAC GAAAGAGAG GCAATGCTG TCTTGAGGA CAGAAAGCT 720  
TGTGTGAC AAGCTTAAAC CACCACTTC ATGAGACATA GATTATGTC TGGAACTAT 780  
TTGACACCG CTTGCGAGTA AACCTTTTA GTGTGTGCA GTGAAACCG TCACTTTCC 840  
CTAAGAGAG GGTGTGTG CAGCGAAAT GGTGATCTG TGTAAACA CAGCTTCAT 900  
CGTATGAT GCTCCTTACT CAAAGAGTG GGTCCCAAC ACCCTTTG AGCTCTCT 960  
TGATTCATG ATGAAACCTG GAACTTTG AGCACTGAT TAACTATAG TCTTAAATA 1020  
ACTTCACA CGTTTCTT AGTTATCT TACATCAGG GTGTGACGA GCTGTTCAT 1080  
AGTCAATTT TGTGGAAT ATTTGAGTG TTATTTGCA CTTAGCCA CTGTGTGAG 1140  
CGTATTTCT TCTAACTCA CCAATATCT GAATAAAGT CAATTTAGG GG 1192

## (2) INFORMATION FOR SEQ ID NO: 202:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 589 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

ATCTTGAGT ATCTTACA GCGAATCTT GCAAGTAT GCTTTCACA AGTAAATATA 60  
GTCTGCTCC AAAGATGAG AGCTGAGTT CTCACGAA TGAATGAAT GAGAAACA 120  
CAGATTTA GATTAAGGA CCGATGCTC GATGTTCA TTCTAAATTT GTTCTGCTG 180  
AAATGATAG TATCTGATG ATTCAGAC AGATGTGTA AGTAAACTG AGTCAGATG 240  
ATGACAAAC AAAGGAGAT GATACAGCA CCGAGATGA CATTAATTT TTAGCCACTG 300  
GTTCAGAGG CAGAGAGAA AGCTAGAG AGATGTTTG TATTAATCT ACTTGATTT 360  
CGGAGATCA GCAATCTCG TCAACAGTA CTCATTTGA GCACTTTCT AGTGTGTAG 420  
ATCAGAGCA AGCTATGAA ATTAAAGAC ACATTCAGA GAGAGGCTT TCAAGCTCTG 480  
AGTGAAGAA ATTCCTGAG AACCTGTG AAGTCTAG AAGAGAACT AAAGAGAAA 540  
ATATGAGAG TGTCCCTTG CACTTCTC TCACTAAC TCACTCCA 589

## (2) INFORMATION FOR SEQ ID NO: 203:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 847 base pairs  
(B) TYPE: nucleic acid

60

- (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

GGAGAGAG CAGCTGCTG GCTCCATCA AGCGTTCCG CAGCTCCCG CTAAACACC 60  
GAACTCTG GAAAGATG AACCTCAG TGGACATCT CAGATTCAC ATGATCTCT 120  
ATGACTTCA GCAATCTG AGCACTCAC ACCGAGCTT GAGAAAGG ATTGACAGC 180  
TGGCGGAAA GCTGATGCC CTGACTGAC TGTATGAGC TGCCTGAGG CAGACAGCT 240  
TCAAGACCC AGCAGACAT CCAATGACT GAACTCAGA GAGAGACA GCTTACTTTC 300  
CCCAATAG AGTGTGAG ACATCTCTG CAACTCTGA CCACTCTGA CAAACAGCT 360  
CAATGAG AGCAAGAG GACTGACTT GAATGATTT GCTTGATAT GCTGTCTGA 420  
GAGAGCTTG GCTAAGTGG GAGGCTTG GCCCACTGA GCGCCAGAT GGAACATGG 480  
TCACTCCAC TGTCAATACC CTATGAAA AACATCTAC TATCTGTGA TGAAGAGCT 540  
CAGCTTCA GTTACATG CAGCGACTG GAGCAGAGC TTGTGATCT CTGGAAGCA 600  
GATTTAGG GCGCCGATC CAGATCTG GAGGCTTCA GTTACTCTG GCGAGCTGA 660  
AGATGAGT ATGAGCTGG GCGCGGCTG GAGTGGGCG CCGTGTGTG GAGAAAGAG 720  
AGAGACCAT TTTCAGAG CTGAGAGAG CAGCTGTG GAGAGAGAA GTTATCTCA 780  
CAAGCTCTG CTTTATCTT TGTAAATAT GTTAAACA GAAAAAAA 840  
AAAAAA 847

## (2) INFORMATION FOR SEQ ID NO: 204:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 852 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

AGAAAGTAC TGCAGAGAG GATGTGATG CTGCCCCAG CATGACCGA AGCTGCTCC 60  
GCCATAGG CGTTCCCGA GATGCGCTG AAACACGGA AGTCCCGGA ACATGTAAC 120  
TCAATGTG ACATTCGA GATGCAATG ATCTGTATG ACTGACGA GATCTTACG 180  
AGCTCAGC GCGCTCTGA GAACAGATT GACAGCTG GCGGAAAGT GATTCCTCTG 240  
ACTGAGCTG TTAGAGCTG CTTGGGCTG AGCGAGCTTC CAGAACCG CAGAGAGTCC 300  
AATGAGTGG ACCCAGAGG GAGAGAGG GATATTTTC CCAATCTGA GGTGTGAGC 360



ATNGCTCTCT TGGCACTCCN TBNACCCAGC CTTGACAAA GCACCTCAGG TGCAGGACC 420  
AAGGGGGCC CTGGCTTGGG GTGGGTGGC TTGCTGATGG CTGCTGGAGG GAGACCTGGC 480  
5 TAAAGTGGGK AGGCTTTGGC CCACCTGAGG CCCCAAGTGG GAACATGATC ACCCCACATC 540  
TGCATACCT CATCAAAAC ACTCTACTA TCTTGTCTATG GACGACCTCC AGCTCTCAT  
TACAAATGCA GGCACCTGA GGCAGACTC CTGGGTCCCT GGGAAAGAGG GTACTAGGGG 600  
660  
10 CCCGATCCA GATTTCTGGG AGGCTTCAAT TACCGCTGGC CGAGCTGAGG AACTGGTAT 720  
GAGCTGGGG CGGGCTTGA GTTGGGCCC CTTGGTTGGA CACAAAGAGG GACACATTT 780  
15 TTCCAGAGCT GCAGAGGCA CTTGGTGGG AGGAAGAGT GTACTCACC AGCCTCTGCT 840  
CTTATCTTTG TA 852

20

## (2) INFORMATION FOR SEQ ID NO: 205:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1354 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

## (3) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

GATTGGGAC GAGCTTGGT GGAGCGAG AGTCTCTTG CGGCTGGG ACTGTGCTG 60  
35 GCAAGACTG GCATGAGCT CATGTGCTG CATCCAGAGA TCTGTGTTT CGGGGGGTG 120  
TGGCTGTCA ATGCCACTG GCACCTTCA GACACATTT GGTGTATCC CATCACTTC 180  
CTGACCTCG CATTGTGCTA GTGTGCGG GGCACCATG GGGCCAGAT GTTTCCTG 240  
40 TGCATGGAG TCATGGGTGT CTGTGACA GCTCTGCTG TGGCTGTGT GGGCCGAG 300  
CTGGATTTA ACAAAGGAGA GAAGCAGTG CACACTTCA TGAATGATAT CCAATATAC 360  
45 AAGAGATGA AGAGTGGC TGGCGAGTG CTACAGAGG CTTGTATCTT CTACAGAT 420  
ACTCCAGGA AGAGTGTCA TGTCTGGG AGCATCAGC GCATCTCT GGGCCGATC 480  
AAGCGTTCC GCGAGTGGG GTTGAACAC GGGAGCTCC GGGAGAGT GAATCCATG 540  
50 GTGGATCTT CCAGATGCA CATGATCTG TATGACCTG AGCAGATCT GAGCAGCTA 600  
CACCGGGCC TTGAGAAACA GATTGACAG CTGGCGGGA AGCTGATGC CTTGACTGAG 660  
CTGCTTAGA CTGGCTTGG GCGAGGCGAG CTTCCAGAAC CCAGACAGA GTCCAGTAG 720  
55 CTGGATCCAC GAGGAGGAC CAGGCTACTT TCCCGATAC TGAAGTGTG GACATCTCT 780  
CTGCCACTC TGAACGAG CTTGAGAAA GACCTCTAG TGCAGGACC AAGGGGGCC 840  
60 CTGGCTTGA GTGGTTGGG TTGCTGATG CTGCTGAGG GAGGCTGG TAAAGTGGGK 900

AGGCTTGGC CCACCTGAG CCCAGGTGG GAACATGATC ACCCCACTC TGCATACCT 960  
5 CATCAAAAC ACTCTACTA TGTGTCTATG GACGACCTCC AGCTCTCAAT TACAAGTGA 1020  
GGCAGTGA GGCAGACTC YTGGTCCCT GGGAAAGAGG GTACTAGGG CCCGATCCA 1080  
GGATTTGGG AGGCTTCAAT TACCGCTGGC CGAGCTGAG AACTGGGTAT GAGCTTGGG 1140  
10 CGGGCTGGA GTTGGGCCC CTTGGTGGG CAACAAAGAG GACACATTT TTCCAGAGCT 1200  
GCAGAGGCA CTTGGTGGG AGGAAGAGT GTACTCACC AGCCTCTGCT CTTATCTTTG 1260  
TAAATATAT TAAAGCCAGA AAAAAATATA AAAAAAATCG AGGGGGGGCC 1320  
15 AGACCATTC TCCATATAGT AAGGGCCNN AUAU 1354

20

## (2) INFORMATION FOR SEQ ID NO: 206:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

## (3) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

TCCCAGGTG CACAGGAGG GCGCTCTGT CTGCGAGCA ATTCAGACT GTGTGGAGC 60  
TCAGCCCTA GNCATTCA AGCTTATG TTGTATCAT ATCTTACGT TTGAGACTT 120  
35 GACTGAAA ACAAAATGT CATACAGTG AATTTATCT TAGAGATCTG TCGAGCTAT 180  
TTCTGTACA AAGTTATAT TGTCTATTA GAGAGTCTT AATGGCTCT GTGATATG 240  
TAACTCAAT TACAGGTGA CTTTATATG CATACAGTA TTGTATGAA GAGCTCAAA 300  
CAATGTGGG ATGTGTGGA AAGTTATCT TCCGCTCTT TCTGTGTC ATTGTCTT 360  
GCGAAGGA TGGCCTGAT GCGAGGAG GCGAGCTGT AATTAATAA TAATTCAC 420  
45 TATGAGACT CAGAGGACT AGACTGGA AGAGCCAG AGAACAAGA ATCCAGCCT 480  
TTCACTTAC AGGTGACAA ACTGTGTA TGCACATGA TGTGTTTTT AAGCTGTAG 540  
CACTTAACA AATGTAAAT TTGCAATAT TAGGAAGTC TGGTGGCAT GAAGAAGC 600  
50 CAGAGGCT TGAATCCAG TGTGTGCC TGTCTAGC AGACACACA GAGCTGGGT 660  
CAGTTGCC TCAGCTGCTT AACAAATTC CTGAGACAGA AAGTCTTAC AAGCTGCT 720  
TCTCGATAC TGAAGGTG AGTTTTGTA ACTGACTGA TTTTATGCA GTTGAATAA 780  
AAAAAGCT ATTCAGAA TTTCAGAG TTCTGAGCA TCTTCTGAT GCTTTACTC 840  
60 CTGAGGCA ATGTTTTAC TTTATGATA ATTCTTGT GCCAGGAT AAGGTGAGA 900

AAAGACACT TTTAATATAT AGCTCTCTT GGAAGAGAC TAAATTGAA AGAAGAACT 960  
GTGACAAATT TCAATATCT ATCTTAAA AACATATC TTAATACAA AAGATCTTT 1020  
TGAATATAG TTAAGACAA TGGACAGC AGTTCTCTT TAAATGATA GTGCTATAC 1080  
TCAATATAT GATTATCAC TACTGCTTT AAAAAAAA ACCAGCAAT TAAATGAAA 1140  
CATGAGACG GATTATAGTG CCTTAACTGA TAAATTTGT GACTTAAAA ATACATTAA 1200  
AATCTCTCT GTGCTCTAGT ACCATCTTA GTGCAATGA TAAATCTAT GTACACTGA 1260  
TCTCTTCTT TAAATTAATA AATTATCAG AGTCAAAA AAAAAAAA AAAAAAAA 1320  
AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1378

20 (2) INFORMATION FOR SEQ ID NO: 207:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1166 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

AAACCACTG AATTAAACC CCTCCCTTC CAGAAAGTT CAGAACCGC CATGATGAC 60  
CCTCAATTTA GATGACCNC AATATTAAG ATGATCTTG GACCCCAAG ACTGACCTT 120  
GAAAGGCGCA CTCAGAAA AATCTTTGA CATTCGAAA CATCTGACC TTGTCAGAA 180  
CATGATGCG GCTCAATGAG AAGCGGACT TCGAGAGAA ACTTACTTAT GATTAGTGG 240  
ATGCTCTGA CCTGTGTGT CATGTCTTC GTGTGTGAT ACTTGACATG ACCGACGAC 300  
GCTCATGCG TGAACAGAT TCCCAAGC TGCAGATG TGTGTGTG AAGAGAGAG 360  
GGGAGCGAGA GCGCGGCTT TGAATGTAT GCTGTCTGA ACTGTAGAA CACTTGGAG 420  
TGAATGTAT CTAAATTTCA ACTGCTCTG TTTTCTGTA CATCTGAG GGGAGCTAG 480  
TGCACAGAC ATGCGCGTG CTTACGAAT GAAAGAGTC CCGGTCTGT CTGTCTGAT 540  
CTGTCTGTA ATGCGGAGG GAAAGATG CTTTGTGCG TTTTCTGTA CAGCTGATG 600  
GTGCGCTGGA AAGTGTCCA CAGTGAAGC TGTGTGAG ACTGTGACN ACCGTGACA 660  
CCTGTGAC ATCGACCTT TGTGCTCTT GATTAGTGG AGCAAAATG GAAAGGAAA 720  
GAAAGAGCT TTTTCTGTA GCAATATAT TAAATATG GTGATATAC ATCTGTGTC 780  
TCTGTGAC CTTGCTCTT GCTTCAATGA CATGATAG ACTGATGCC AATCTGTG 840  
ACCAATGCT GAAAGAGCT ACTTAGGGA AAGCAAGAT AGTCTATG CCAATACAA 900  
GTAAATTTT CCAATACAA GGTGCTGT TCTTTTCT AAGTGAAG CATGTATAT 960

TCCCGAGAT GTTCACTCT GACCTATG ACATATTTT TGTATGTAT AATATAGTG 1020  
TTGGGCTTT AAGTGTGAA GACTGTGCG TTGGGCACT GCGCGCAGG GTTGTGCG 1080  
GGCTTTGCG CCTTGAAGA GTACCTTTA GGGCGAAGA TTTGTGAT TTGCCATTA 1140  
CATCTTGA CTGAGGCTT TTAATA 1166

10 (2) INFORMATION FOR SEQ ID NO: 208:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 697 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

TACTTGAAG ATTATAGAA ATTACATAT AGATGACAT TCAATTTAG AAGAAATTA 60  
GTGCTTTCA GTGCTTTTA TTTGATCTT GAGAGAGCA GACTGCGAS AACATGAC 120  
CCAGGCTG ATATGAGAT AATCTTGA GCGAGAGCC AGCAAAAC AGCAATCTA 180  
GAAATTTGA ATTGAAGAT TTCTGCGAG CTGCGGAAT GACATGCAA AGCTATGCC 240  
AAGAACTGC AAGATATTC TGTCTATAC TGTCTTACC ACCGATTC AGCTGCCCA 300  
ATTACATAT GAGATGACT AATCTCTTT ACTTTATCA TTAAGTGA AGCATGAC 360  
AAAACTTC GACTTCCA TATTAAGGCT GTGTCTCT GTGTGCTT GAAATAGAG 420  
CATGACAT ATCTGAGAC ATGCAATTA TCAATATAT TCACTCTC CCGAGACTC 480  
CTGATTTGA AATTCAGAG TTCTGAGAT GATCTATAC ATCTAAGAT TTGAGAGCA 540  
TTGATTAAG TTAATGAT AAGAGAGAT TGAATAGGA TGTGTGCTA TCTTCTACA 600  
GTTCCTCT TACAGAAA CTGAGAGTC AGTTATGAG AATATGAT AATTATGCT 660  
AAACATTA AATATTA AATCTGAG GGGGCGC 697

45 (2) INFORMATION FOR SEQ ID NO: 209:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

50 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

CTGATGAC CTCTATAG TGGCGTGC CAGGCGCGG GTGACCTTC GAAAGCTTC 60

5 GCTGCCAGAA ACCATGTTCA AGGTAAATTAA AAGGTGCGTG GGGCGAGCCA GCGTGAAGCTT 120  
GCTCAAGCTTC AAGGCTCTATG CAGGACGAAA AAGGAGCTCA CCGTCCGAAA ATTGCGGTGAA 180  
GGTTGATGAG CTATTCACTGT ACTGAGTTCC TGAGGCTCAA TCGAAGTATG TCGAGGAGCC 240  
AAGGAGCCAG CTTGAGGAAA GCACTCTACA GCTCCGACAC TATTGCGAGC CATACAGCAC 300  
CTGGGTGTCAG GAAGGCTACT OCCAACTTAA GCGGAGATG CAAGGTTTGG TTCAATGCGG 360  
GTTAGACAGC TATGACTATC TCCAAATGCG ACCTGCTGGA TTTTITGCGA GACTTGCTGT 420  
TATTGGTTTT GCTGGCGTTA TTGAGCTCTT TTGGCTAGA GGTTCAGAAA TAAGAGAGCT 480  
AGGTATGCG CCGGTGTTCA TGGATATAG TGGCTGCGTC TATTATGCGC AACAAAGCCT 540  
CGGTGTTGCC CAGGTCAGTG GGGAGGANTT ATATGACTCG GGTTHACGAG GATATATAGT 600  
CATAGAAGAT TTGTGAGG AGAAGTTTCA AAGCCAGGA AATGTGAGGA ATTCACCTGG 660  
AACTAATGAG AAATCTCAT GTTCTGCCAT CTTAATCAGT TATGTGTAAA CATTTGGAAC 720  
TCCATAGAT AAATCAGTAT TTCTACAGAA AATGCGGATA GAGTGCAGTA TTGATGTAT 780  
TAAATTTGCT TTCTCTGTTCA GGAATACTA GACCAGACT CTGTATATCT CTGTGAATC 840  
ATCTAGAG CAAGTAACC TGGATGCGT TCACTAGAG ATAAATGACA AGCCTTAGAA 900  
CTGCTGATTC TCAATGTTCT ATTATGTAC CT 932

(2) INFORMATION FOR SEQ ID NO: 210:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

45 GTCATCTTTT AATAAAGAG TTCTGTGTTT AAGCTTTTC AAGGAGCAG ACCACCTTCA 60  
AGATTCGCC TAGGCTTGAT ATGTGCTTAA TTCAITTTAT AAAATTTTCT CTGTCTCTCA 120  
TTTAAAGCT TTGCGTATAT AGTCAGAAAT GTGCTAATA ACAATATTT TTGTATTTAA 180  
TTTAGGAGG ACTAAGAGGA AGAATAATCA AAATCTAGTC TTTATGTAG CTCGAGCAT 240  
ATTAGGCGTT AAGGCGCTTT TCTAGTTTAA TGAAGATTTG TACTACTGAT TTTTATAT 300  
TCCGTGTTTT GAGATGACA GATCTCTGGG GAATTCGTTG AGTTACAGT GCATTTCCT 360  
GTGATGCTTC TCAAGCTCAG ATCAGTTCTA TAACCCAATG ACAAGCTGTC TCTTTGTTTT 420  
ACTGTCTGCT GAATGTGAG CTCAGTTTTC CCAGAGTGG TGTGTTTTAG ATGAGTCA 480

5 GTGCTTTTCC TCGGTGGGAC AGTTGCTGGC CCGTTAAAT TTGGGTATG TCGTTCCAG 540  
TATCTAAAC TCGAGTCTGA TGTGTATATG CTATCTTAC TTTTAATTT ATTAATGAT 600  
5 ATGTTGATTA TCTTCTTGA AGGTTCAAC TTTTCAATTT GATAGAAATA AGTTTTTTT 660  
C 661

(2) INFORMATION FOR SEQ ID NO: 211:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 592 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

60 GAACTGACA TTGTTAACA CACTAACA GAAGTACTTA CCGTTGAG ATTAAATATA  
120 TAAATGTGA CATGATACAT GTACATGAT GGAATGACCA GATGCTTATG GTCTACATTT  
180 TCGTTATCC TGTATGATTT ACCTTCTTA ATCTTGTTTC CTTAAGTGC TAAATTCCTC  
240 TTGAGTGTTT ATTTCTAGT GACAGATGC TACATTTCT TACACCTGG CAGAGGGAG  
300 AGAATGTTT TTGCGGTGG GTAACTAAT TTTTCAAGTA AATATCAATA GATCAGATG  
360 GAAAGGGA GACAAAGA GTTATACAA AAAACAATG GTTTTTTAG CCAATTTGACT  
420 GCGCTTTAA ATAGCTACA AGCAATTCAC GTTAAATAC ACTTTAGTG AAATAAATG  
480 TCCCACTATA GTATGTGCTT CAAGAGGCA ATGTGCTTT AGTGCCCTTA GCGTAATTT  
540 TGGTCATTTG ACATCAGAGA TGTGTAAAT ATTCACCTTA ATAGCACTT ATTTCTCAAT  
592 AGTGTATTTT TTTTGGCTAG CATTTNCTTT ACCACTAAC TTGTTGGATA GC

(2) INFORMATION FOR SEQ ID NO: 212:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 938 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

55 TCGAGTGCT TTCCAGCTCA ATGAATCTTA TGTCTCCCT GCAGGTGTTT GGTTTTCAAT 60  
GTCTTCTCA ATTTTTTCC TATTTGCTT TGGAGTTTN CTTTGTGTC TCGTGTGTTT 120  
GCCAGCTTT AATAAACA GCGGACACA AAACCATAG CATCTGAA CAATAGGGGG 180

	CCGACATTCG ACCGACATTCG TCACCTTAAAT GACATTCAG AAAAATATCG AATGGAAA	240
	TCACATCAG AATATTAAT CCGACATTT TATGCAATAT AATGATGAT TTTCTAAAT	300
5	TTGTTTCTCG TGGGAAATG TGGCTTCCA ATTAAATCG CATTCTCTC TTGCAAACTT	360
	TTTGTTTTGA CTGATTAAT TAAAGTTTG GAAAGATTA TAAATTTGAG AGAGTTTTC	420
10	AACGACGGA TCACAAAGG TCTGATGAT AATCTGTTC ATGTCTTTT ACAGCAACT	480
	ACATTTAAG ATGATTAAT TCACAAATT AATATCTGT GTATGTGCT CTACATATA	540
	AATGACATCG CTCACATTA TGGGTTGCTG TCCATCTCG CAATATCTG CCAATCTCT	600
15	TTATGACAG CACGAGAAA CCAATGATG GTCTGCTTT CAGAAAATC CTCGATCTT	660
	TTCTGAAAC TTATTTTCT AATATCTGT TTTCTTTGA TTTGTTTAC CTCACAGAC	720
20	CATTGTGAC AATGATGAC CTCATTTCA TCAATGATC CTGTGTTTA ATGTAATA	780
	CATTTTCAAT GAGATCTCG ATGACTTCT ACGTTTAT TTTTCTGTA AGCTCAATG	840
	CTGAAACCA AACAGGCTT TTAAAAACT GTTTAGAGA AAACAAAA ATCTGTGTG	900
25	GTGTCTCTT CCGTTCAAA CTCATTAAA ATTCCTT	938
30	(2) INFORMATION FOR SEQ ID NO: 213:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1079 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
35	(D) TOPOLOGY: linear	
	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 213:	
40	AGCTTCGCG GAGATGTGT GCACTTAAA GGTATGTGT GCACTGTGT TGGGGAGAT	60
	GGAAGCTTT TTAACTGTT GCGCCCTCTC CTGTGCTGAC GTGGGCAATC CCGGGGAGAG	120
	TGAAAGCGAG GCGCTCTCT ACCTTCCGAG TCCAGCCAGC CTGGGCGCGG GCGCGCGCCC	180
45	CGAGACATCC GAGGATCTCG TTCTCTCTCG GTTACGTGGA CTGTGAGCT GTCTCTTGT	240
	GCTCTGAGCC CTGTGCGAGG TTGAAAGCTA CCTGCGAGG TCCACACAG GCGCTGAGGA	300
50	CGAGAGGAA GGCATTAAC CCACTTTGAG GAATCGATG TCCAACTCT ACATCTGAGG	360
	ATTCATCTCG GAACTGTGT GGCATTCGAC TGGGGCAGAT GTCTTTGAG TTCTACAGA	420
	AGAAAGATC TGCCTGACCA TTCTACAGC AGTCGATCC ATGGGAGTGG TGGAGCTGA	480
55	AGCTGCAATG GTTACCTCTG GCGACGAGC AGAGCGGCA GATCTGCGG GAGAGAGTGG	540
	GTGAGAACT CTGCGAGAGG ATCATCTACA TCTGTGAGGT GATGAACTGG CATGATGACT	600
60	TGCCCAAGAT GCGGACAGAG TGGAGGTGG ATTAAGTATT TCACACAGCC TTGCGGAGAG	660

	TGCAGCTTA CCTGTACAG ATCTCTTCC AGATCACTGA TGCCTGAGC ACCTCACTGA	720
	CGACGCAAT GCGAGGCTTC ATCAAAAGA CCGTGGCTT CTGAGCTGCG CTGATCTCT	780
5	GGAAGCTCT TGAATGCTCC CAGAGCTTGG CTTTTGGGAA TTGCATCTTT GCGCTTTGG	840
	GCTGTGAC CTGCTGTGG TCAATGTGA GACTTGGAG GGGCAGCGCC CCGTGGCTTC	900
10	TGGTTTGT GATTGCGAG CTCAGCTCAT CCTTTAATC TTTGCTGAG GTTCAAGCTT	960
	GCTGTATG TCTGTCAATA GCGCTGTGAG GATTCGCTCT CTTTCTCAC TGTACAGAG	1020
15	AACGACACT GGAATCGGGA ATTAAATGTA GAACATGAGT TTAGCTGAAA AAAAAAAAA	1079
20	(2) INFORMATION FOR SEQ ID NO: 214:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 3791 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
25	(D) TOPOLOGY: linear	
	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 214:	
30	TGAAAGAGC GCTCTTGTCT GCGCGCGGCG CCGTGCATC CTGTGAGAA GCGCGCGCGG	60
	AGCCACATC ATGCTGTGAG ACTTACAGGA AGGCTTTGCG TCGTGTGTA CCAAGCAAT	120
	CGACGATTA TTTCAGCAG AATGCAACC CTTCGAGTGG CTGAAAGAG CAGAGACAA	180
35	GAAAAAAGA GCGCGCGGCG GCGCGGTGG GCGCGCTTGG GCGAAAGCG CATGAGGCGC	240
	GCGCGCGAGA CCAACTGCA GCGCGAGCG AAAGCTGCG GCAAGAGTC CAGAAAGAC	300
40	GCGAAGATC GCTGCGCCC CAGCTTTGCG GTGTGTGCA AGAAAGAGA GACGAGCGG	360
	CGCTGCGGC TTTAAGAAAG AAGGATTAAG AGGATTTGGA AGAAGCTTG ATTCAGAACT	420
	TGAGGTGAA GGGAAATTA TTGATTAAGG ACCAGAAAG CAGACACTTC GTGAGAGAG	480
45	ATTCGAAAG CCACTTGAAG AAAAGGTGTA AGAGGCGAA TTTTCAGTTG ATTAAGCAAT	540
	TATTAACGA CCTATTGAGG GTCTGTGAG TCTTGAAGA GTTCAGAGGG GCGCTGAGCG	600
	TGGAAATGCG CAGAGAGTGG GATTTGATTC TGTGTGAAA CTGAAATTTG ATAGCAATAG	660
50	TGGAAATGAT AATCTTCTTT TTTCACATTA CAGTGGCTTG AACAGAGAG ACAAGCTGCG	720
	AGTTAGGGA TCTGACACT GCGGAACTCT CAAAGAGAA TTAACTGACT TGAATCATC	780
55	AAATGTACT GAGGAAAGC CTGAGGTGA AGAATCATC CAGTGTGAG ACACTGAAA	840
	TAAAGAAAT GAAATTTGAG AGTTAAAGA GAGGCTTCCA AAAGAGTGA CTTTGGATGA	900
60	GTGAAAGCT ATTCAAAATA AGGACCGGCG AAAAGTGAAG TTAAATATTC GAAACGAAA	960

1020 TGAAGTCTCT GATGGGAGT GAGAGAGGG ATTGTCTCTT CATTAATCAA AGAGTGAGA  
1080 GGCTCATCT GAAGTTCCG TTATGGACCA TCATTTCCGG AACCCAGCAA ATGATATAC  
1140 GTCTCAGCTG GAGATCAATT TTGGAGAGCT TGGCCGCCCA GAGCTGGCG GAGGGGAGG  
1200 ACGAGTGGGA CGTGGCGGTG GTGGGCGCCC AAACCGTGGC AGCAGGAGCG ACAAAGTCAAG  
1260 TGGTCTCTCT CCTGATGTGG ATGACCCAGA GGCATTTCOA CCTCTGGCTT AACTGGATGC  
1320 CATAGACAA CCGTGGTCC TTGTGAGCC GTTCTGTCA AGCTTTTTC ATCTTTAAGG  
1380 ATTCCAAAG ACTAAGAAT TAAAAAAMA AAGACTGTCA TTCTATACAT TCACACCTAA  
1440 AGACTGAAT TTATCTGTCT TAAAAATGA CTCTCCCGC TACACAGAG TAACAATAT  
1500 GGTAGTCACT TTGTATTTA GAATGTATTT GGTAGCAGGG ATGTTTTCAT ATTTTTCAGA  
1560 GATTATGAT TCATTCAAA TACTTTTGTG TTGCTGCTTG CAATATGCA TTTCACACT  
1620 TGAATATAG GTGTGACAG TGTGTACAG TTAAAGCTT TCATCTCAT TGTGTTTT  
1680 AATTAAGAT TTAGAGTTC CCCCATTTAC AAATGTGTTT TAATATTTGG ACATCTGCT  
1740 TTTAATACCT GCTTTGCATA TTACACATG GTCAACTCGG ACATGTAAA CTTCATTTG  
1800 TCAATTTTA TGTGTGTGG AATCTACT ATATGTATTT TAACTTAGTT TTAATATTT  
1860 CATTTTGGG GAAATATCTT TTTCATCTC TCATGATAGC TGTATATAT ATATCTTAA  
1920 TCTTTATATA CAGAAATATC AGTACTTCAA CAATTTCAA GCATATTCG TTATTAACG  
1980 CTTCCTCTCT GCATGCTCA TTAGTTTCAA ATTAATACG ATTTACATTT TCAGCTATAT  
2040 TTACTTTTTA ATGCTGTGAG TTCCCATTTT TAAATCTTAA ACTAGATC TTATTTGGT  
2100 AAGTGTGTT AACTACTTAA TTGTGTGGG GCACATGCG TCAGTGTAG TAGTTTATA  
2160 GGTATGGGT TTTCCTCCC CTTCACCAG GTGGSTGGA TAAGTTCAT TTGCCAATGT  
2220 GTATATTTTA ACTGTCTCG TAAATTAAGT GTCTGGCCAT TTGGTATGAT TTCTGTGTGT  
2280 GAAAGTCCC AAATCAAAA TGTATATCC ATATATCAGC ACATTTTAC CTTCTCTGT  
2340 TCTAAACAAA AAACCAAGG GCGCTGTGG GTAGGGTGA GTGGGGAGT ATTTATTTT  
2400 TTGGATTTG GAGACAGAC AGCTTTACT TTGAAGTTG GAACAGCAGC ACTATACATG  
2460 AATATTAAC CAATAACCTT TACTGTCTCT AATTTCTTA GATCTCATT ATTTGTGTGT  
2520 AAGTTGATTA TTCCACAGA AGTGATATTT ATCTCTTCTC TCTTCTCCA TTAGAAATTT  
2580 AGGTAAATTA TGGATTTCTA TAAITGGAGC ATCAACCACTT ATTAACACAC ACATAGATG  
2640 ATCAATTTAA AAGTTTCTCT AGATGTCTT TTATTTCTGC CACATTTAT GATTAAGAT  
2700 GAAGGAATTT TTAATAATTT TTATAGATTT GTTTGTACAG TCAATTTTAT AATGTCTA  
2760 CCTGTATATG GTATATGCA GTTTTAAAA TATTTGACAT CTTCATCTTT AAACATTTCT

2820 ATTAGTCTGA TTGTTCTCA CATATACTTC TAAAGAAAC TTATATGTTA TAAGAGTTAC  
2880 TTTTTGCATA AGATTATTA ATCTCAGTTA CCTACTATTC TGACATTTTA GAGAGAGGT  
2940 AATTTGTTTT AATGATGAT AACTGTGTGC TGTGTTTTTG GATCTTATGA TGTGTGACAT  
3000 GTTCTGCACT GGTGCTAAG TCTATATATA TTATATATTT ACACATATAC GTCTTACCCA  
3060 GAGATTAAAT TAGTCCATAT GAATATTTGA CCCATTTGTC ATTGAGACAG CAACATAGC  
3120 ACTCTTAAT CAGTGTGTTT AGACTTTTCA AGTATCTAAC TCAATTTCCA ACATGTACCA  
3180 TGTTTTATA ACCTCTTGAT TTCCAGCAAC ATACTATAGA AAACACCTGC TACTCCAAAC  
3240 ACACCTTCTC AGTGTCAATC ATTGTGTGTG TGAAGACAA CATAGCATA TCTGTATGT  
3300 TCCAGCTTT CAAGATAGCC TGAATTAAT AAGTTGTGTC ATTATTTGTA TCTGATGAT  
3360 ATAAATTTGC CTCTTATGTC ACTTGTGTGC AAGAGCTAAA ACTGTGAGC TAACTTTCTC  
3420 TTATTTGTGG GTATTAAGT AAAATTAAGA TTATTTTCCA TCTCTACTTC TTAATAAGTCA  
3480 TAAAAAAT CAATATAGAT CATGTTTAT GTCAATGTGT TCTGTGKTC TGAATGTGT  
3540 GCACACCTCT GTGTGTTTAT ATTTTATTA TTGATTTTAA TATGGGTTTT TTATTTCTTA  
3600 AAAACAGCG TGTGTAAATCA CATTTGGGA GGTACTTAT CTTAATGACT AATGACTTAA  
3660 TTGGAAAGT TGAATTTCTG TAAATATCAA AATCCAGGA CTCTTTGGA TTAAATCTAA  
3720 TTGTCTCTC NTAGGCGA TNCATTTT TTGATATGAG AAGTTTACG ATTCGAGAT  
3780 CTACTTTTGG TTGACAAAGC GGCTATATG TCCGGGGGA AATCCCTAAC NGTAAAGAT  
3791 CCCAATATG G

(2) INFORMATION FOR SEQ ID NO: 215:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1334 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

	GGTGTCTGGG AGAGAGCTGT CCAAGGAC CAGAGAGAC TCGATGACA TCGACACAA	420
	AAACAGGATC ACCCTGACGA GGTGCGGAG ACAGTTTGAC AACTTTAAC GGGTCTTCAA	480
5	GGTGTGAG GAAATGCGG GGTGCTGGT GAGCAATAT GAGCAAGCT TGTCTCTCTC	540
	TGAGCGGTTC GCGAGGGACT ATGAGAGAT GTTCTCTTT GTTACAGAC GCTTTGAGAC	600
10	AGGAAAGAAA AATCTGAGT ATCTGAGCTT CGGTGACTTT GCTTCTGCG CTGAGCTAAT	660
	GATTCAGAAC TGAACCTTG GAGCGCTGCA CTGACAGATG GATGACATGG ACATGAGCTT	720
15	AGACAGGAAA TTTCTTCAGAG ACTTGAAAGGA GCTGAAAGTG CTAGTGGCTTG ACAGAGACT	780
	TCTGAGCTTG CAGAGAGAGC TGTGTGAC TGTCTCGCG GAAAGCTGG GGTCTCTCTC	840
20	TGAGATGAAA GCGCACTTCA AAGAACTCTC CGGAGGAGCTG GTTAAAGTGG CGGCAAGCT	900
	GACCCAGAT AAGAGTGTCA GAGAGCTGTT TGTGAGATC GTTGAAGAT TTTGTGAAAC	960
	CTGCGCGCTC GACAGCTGAC GACTAGGGA GTTGCAGTTC TTCTTGATC AGTATTCAGC	1020
25	GTCTGTCAAC TCCCTGATG GCTTCCAGCA CGAGAGCTCT GGAACAGTAA GATGAGACCT	1080
	CTGCGCGGCT GCTCTCTGCG CATTGTATAT GACTGAGATG GTTCCAGAGG CTGCGCGCAC	1140
30	GCTGAAATA AAGTTGCTCT GAGTTTGAG ACTGTATCTC GCTTCCGAGGA GAAATGCGG	1200
	GCGGTGCAAA TGTGCTGTG TGTGTCTCTG AGCAGCTGAT GTTGGTGTAC AAGATGAGAT	1260
	GTGTGATGAT GCTTCTGTGG AACTGAGACA TATCTGAGAG AATGTGTCT GTCTGAGGCC	1320
35	CATTCACGAG AAGA	1334
40	(2) INFORMATION FOR SEQ ID NO: 216:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1511 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
45	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 216:	
	GTGCGCGGGA TGTGCGAGAG GGTGTCTCTG CCGCAGCGCG GCGCGCTGCC TACCTTCAG	60
50	ACTGTGCGCT AATGCTGCAA GCTGTGTAC CCGCAGCTCT GTTGTATGCA CTTTGAGCG	120
	CAGAGGTGTA TGGCTGTGAC TGAATATATC CCGCGGAGAC GAGCGATTTA CCGCATGATC	180
55	CTGCGCATCT CCGCAGCGC CCGCAGAGAG GAGATAGGCC TCAATAGGCT TTTCCGCGCG	240
	GAGATGAGAG CAGTTTTCGA GAGAGACGAA ATGATAGCGG TCTGCCAGAA TGTGCTCTTG	300
60	AGTGCAGAG AGAGCTTCT TATGTGAC GAGCTGCGAA AACAGAGAT CTTATGAGAG	360

	ATCTTCCGCA ACAGGTTCTT GAGCGCTTTC CTGAGAGGAT CAGATGACAA AATGTCTCTG	420
	CGCTTTTTC TGGCGACAAA CATCTGCTTG GTCAATGAGAG AGCGCAAGCT CAGAGAGAGC	480
5	GTACAGATCT TAAAGACTGT GCAATCTCTG CGCTGTCTTG GTGCTGCAAT TGAATGACAC	540
	ATTCCTACAA GCGAGGCTTT TATCAATAC TCCAGAGCTCC CGAGCTGCC GCTGTGTGAG	600
10	GCGAGACTTC TGAAGAGCTT GAGTGCCTTC ACAGCGCAGA CCGAGCTCTT GCTTGCAGAC	660
	GAGCGCTTCC AGCTGACAC CGTGTGTGAC GATGATGATCA GAGAGCAGAG CAGAGAGGAT	720
	TCTGTATGT GCGCGAATGG GAGCCAGAT CCGTACACTG TTGCGGACTC GTAGCGAGGCC	780
15	TGTTTACGCA GCGTGCAGCA TAAATGACAT CTGCGTTATT GGTGTGCTCT TCCCTCAATGG	840
	GACATGTGAA AAGACTTGGG GTGCGGAGAT GTGTGTGTCA CTGCGTTTTC ACTGATATAG	900
20	AATATGTGAG GTTAAAGGCGC ACTGTGAGAT GCGAGAGGAT CAGTTTCAG TGTGATGTAC	960
	GAGCTTCTCT CTTAGTTTTC CCAACTTGGG ACTGTATAG AGCAAGCTCT CTCCATTTCTC	1020
	GAGTGCAGG GCGAGAGATC TGAAGAGATA GAGCTATGT CCGCTGCTCT CTTGTGACT	1080
25	GCTCTGCTT GCGAGGAGCTC CTGAGCGAC CGTTTGTGGC AAGAGCTGCC ACTGCGAGAG	1140
	TGAGTCAAC AAGCAGTTGT GCTGAGATG GCACTGTGTG AGAGCTGCT GTGTGCGAG	1200
30	CTTTGTCTTG AATGTGTAC ATGATATAT TCTTTTACTG CTGACAGAT TGTACCCATTT	1260
	TGACAGGAAA GAGAGGAGAA AATTAATGGG CTGTCTGAG GTATATGAT TGAATAGTGG	1320
	CAGAGAGAG ACTTGAGCAA AGCGCTTGC TCTGAGAGAC GGTGTGAAA TTTCTTCACT	1380
35	AGAGCTTCTT GATGAGGTTA CCGAGAGTGG GGTCCGATCC AGCATTCAGG TGTGCTTCAA	1440
	TGTATATCT CAGCGCTGCA GGTGTACCT GTTAAAGTT TGGAGAGCT GCTTTATAT	1500
40	AAATGAAAT A	1511
45	(2) INFORMATION FOR SEQ ID NO: 217:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 642 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 217:	
	AGGCTTACT TTTCTTCCA CAAAGATTC GAGGCGAGCG TACATGTGAC TTGCGCACTGT	60
55	GAGAGTTTC ACTTGAGAGG TCTAGGAAA GACTGGGCAA TTGAGCAGAG GAGAGCGAGC	120
	TGTGATCTTG ACCATGAGGC GAGCGCTTTC ACCTTGAGCTG GCTGTGCTCT GATCTTGAAG	180
60	TTTGTGAGG TTTCTCTTCT TTGATTCCT CAGCTAGTGG ATTAAGACTG GAGGGGAGAG	240

445

ACCGCCCTTG GACGTGTTTC TTAACTCTA TCCATATAT AGGCGCGTGG GATGGTTGTA 300  
GAGTAAGC AGGATGTCG TGTTTTAA GACAGCTTG GACACGGCC TCTACACCT 360  
AATTTTCT CTCTGTAGT GACAAAGGT GTAAATTAGC TTACAAAGG AACAGCTGC 420  
CGTCAGCCAG AGTTCTGAAG GCAATGCTTT CAGTTTCCCT TGTGACAT TCTCTCCAG 480  
TTCTATGAA AGCAGAGAG CTTAGGGGGC CTGGCCAGAG AACACACCA TCTTAGGCTT 540  
GACCTGTGAA CAGCAGGGGG TTGTGTCTCT GTTCTGTGTT TCTGCTTCCC GACCTTCTC 600  
AATAACCTT ATTTCCTATT TTATATTATC GTNGGTGCTG GG 642

(2) INFORMATION FOR SEQ ID NO: 218:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1241 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

GTGCCACTG TTCCATTTTA TCCATATAGA TTCCATTCTA GGGCCAGCC GTCTCTGAC 60  
TGATGGTGT CCCTTAACC CTGGCAGGT ATATAGAAAT TTGGGTGAAT GAAGAACCC 120  
AAATAGGCA GATAGTCCC CCAGGCCCTG ATATCATPAA AAGGCTTGGG AATGCATAT 180  
GTAAATGTCC TTATGCTTTT TGTGTGTTTA GAAAAAATA ACAGATGGG CTCAGATGA 240  
TGCTACGTA AAAATGGTTC CTAGCTGTGT ACTCATAACT TTCTTTTGA TTGAGTAGTG 300  
AAGGAGGA GGAGGAAGG AATTAAATG TCTTCTAGT ATTCTCTGA CTCAGTCTG 360  
ACATATGGA TAATTAACCTA TATTGAATG CCAGAAATG TATCTGAAC AAGGACAG 420  
TTTGACAT TTATCATCCC TTATATTAC ATATTAACCT AATCCATTA ATAAACATAT 480  
GAATATCCA TTGACAGG CAAGGACCC TAACCTTTT GTTCTTTT CTACATAGA 540  
GAATGATTT TTTTGTAT TTCTTTAGG GAACCTATAT AATATGACC CAGTGAATC 600  
TTTGTGTAC TTAACTTAT GAATCAGGT TACAATTGAG TTGATCTAG ATGTTACTA 660  
CCTTGAAGG GATGTGTGT CCTTATGTA CAGAGGCCAG AGCCTCTCG GAATAAACA 720  
AGCAGTTCA TCCACACACC AACTCTAGC TTTAGTGCCA GATGGAGTG GTCACAGCT 780  
CCCAAAATG GGGGCTTTGG ATTTCCACAC CATCCAGCT GTGTCTATC TTCTCTTTC 840  
ACACTCTGA TGTAAATTTG AAAATGTGCA AATCAGCTCT GAATTCCT ATAGCATGAG 900  
CAGCTCTTA TGACACATA ACMAATAGTT CATATGTGA ATATAGAA CTCCTACAGC 960

446

CTCAGTTAC CATAAATTTT CATGTTTGTG GAATGTATAT TGAATATACA GGGCTAAGA 1020  
ATTACTGCA AGTTTAGCC TGTGGTAAAT ACCTTAGGCT TATTTAAATA TTGTATATTT 1080  
5 TATTAAATG TTCAATGATG TTTCGAAGA ACMAATATAT CAGGATGCC TCTTCCCAT 1140  
GGGTCTATTT TTCACTCTCT TTCTGTAAAG AAAAAAGAAC AATGCTTTAA TGTATTTTAA 1200  
AAGTTTGTGG TATAGTTTCT AATTCCAATT TTAATAAAG T 1241

(2) INFORMATION FOR SEQ ID NO: 219:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

TGTTATGTC ACTTAAGACA TACACATG CACACACA TACATATCCA TTCATCATTT 60  
CATCAAGTG GTGTTTCCAG TGTCTGTG TGACTGTGTTA TCGATTTC ATTTCOAGT 120  
GAATTATGAG TCGAGGGCA CTTTCTAAC CAGATTGTCT TTTCAGAAC AAGACGCGG 180  
RATTAGGAA GAGTTTGAA AGAGGAGAG GCAAGGAAG AGAGCTTTAA ATTGAAGGT 240  
TAATTTCTA AGAGAACCT GGGCTGAATG ACTACAGTGT TATACCCCTC AATCTTTGCA 300  
GGTGGCAGT GACACTGCT TGTATCACTC TGTCCAGGT ATAAATCCAT ATATCCAAA 360  
AAACACAT CCATCATCA ACATATACAT GGTTCGGAT CAGCAGTCA ATAGTTTGA 420  
GAGGAGTTT GTTCTTTT TTCTCTATT ATACTCTTAA ATGTGTGCA GTTATCAAAC 480  
AAGCAAGAG AAAATTTGTT TCGAAGAAC CTTCATAGC CCTTTCTAT CAGTGTCTTT 540  
AAATATAGA CTAATACAC ACATCTGCC AGTTTTTCT TACAGTGCA GTATCTTAC 600  
CTCCATTTA ATATTAGCCT GGTATTTTTC TCAGTATAT TTACCTGTGA CTGTATTTG 660  
TTATTAAC AGAAGAAAA ACATCCAAA AAGAAAAAT TACGTAGC GCTCTATAT 720  
ACTATATAT TATATATAT ATGTGACAT TTTCGAATAC TGTGAAGTT TATCTCTGC 780  
ATATACTTTA TACGAGTAA TTAGGCTTTA AAAATAGGAA AATAATTTT ACAAGTTTC 840  
TGTTTGTGT GGAAGAGTAA TTGATGTGC TAAGATGAT GTTTGTTTTT TCGGGTTTT 900  
TGTGTTTTT TTTTAAATG TTACAGCAC TTTTGTGTA AGTTTCACT TCCGAGGTAT 960  
TGTACAGTT CACACTGTT GTGAGTTTG AATATGAGG AATAATTAA AAAAAAAA 1020  
AAACCGCG GGGGCCCCG TCCATTTG CCAAGGGG CGGTACGGG GTCACGGCG 1080

(2) INFORMATION FOR SEQ ID NO: 220:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1258 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

10 TGAATGAGG GCTTAAGAT AACATATG GATTGAGT GTGTGCAAT AGGTTTAC 60  
15 TCCATTTG ATTGAGTT ATCCATTA ATTTTTACA GTGAATTT ATTAAATAT 120  
20 ATTATGATA TATTTCAGT GGAATTTCT CTGAAGGTC TCAATGATC TGAATGAG 180  
25 ATATGAGGC TTCACTGTG GGAATTTCA GGGAGGAGC GCTTCACTC TATGACACA 240  
30 TTGTATTAT GGAATGCTC TCCCTGTG ATTATGTTG AGTTTACAA TCCCACTAC 300  
35 TTACACACA GCGAAGGTG GAAACAGAC CTACACACA AGCTACACT ACCCAATCA 360  
40 GAAGGATGC CTTGCTGCT CTGGCCAC AGTTGATC TGTCCCTTG GCGATGAGC 420  
45 GGGATGCAA TTGACGGTT CAGTAAGAG AACGTTCA CAGTTGAGC AGAAAGCTA 480  
50 GTCAAGACA ACAAATAT TATAGAGCT ATGAGATCC TCAATTGAA GATGATGAA 540  
55 AATTCACAG AAGATATAT GTCTTTTCC ACCCAAGGG ACTACATCA TCTGAAAC 600  
60 AAGCTTCCA GCTGCTCTG CTGCTATAG TGTTTGGTT ATTTCATC CCAATCTCG 660  
65 GAGCTCTTT AAGCTCTC GCTTGGTGG CCGACATAC TATTTTATA AGTAAATTT 720  
70 AATGTCTCC TGAATCTGT CAGTAAGAA GAGCAATGT CACTTGAAG AAGAACCTG 780  
75 AACCCAGTG CATTCTGCA TCTCTGAT TACCTTTTA CATTGCTCT TCTACATTA 840  
80 GTTCCATTA GTCCCTTGG TGTAAATCT TCTATAGC CTTCAATTT TATATCGAA 900  
85 TTTTGTGAA AGATATGAA ATCAGCACT GCGTTTGA GATCATATT CTCACTACT 960  
90 TCTGAGTTA TTTTTCAT TGAATTCAT TGAATATG ACTTCATTT GAGAGAAA 1020  
95 TGAATGAA TGTGATTTCC CAATTTCT GTAGCGGTT GTTTCATATT CTTTGTCT 1080  
100 TGAATGTA ACATGATTT CTGATGCA GAAAGAGGG TCTGGGATC TGTGATTT 1140  
105 TGGCTACAG AAGTCTCCA GAGTCACTG TATTTTGA ACTTCTACG TCAATTTTA 1200  
110 GTTCTCTG TCTTGGCAT CAGATATAT TCAATTTTT TGGCCCGGG CAGGCTG 1258

(2) INFORMATION FOR SEQ ID NO: 221:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1693 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

10 CAGATATAT GAATATGAC CCTTAAAAA AGAGAAAAA AAAATGAGC GGTCAACT 60  
15 AAGACAACT TGTCTATTA AAGATATTT TATTTCATA GAAAAATTT AATTCAGG 120  
20 ACTATTAAT ACTGATTAAC TAGAAGTTC TTTTAAATT GACCTTAAA ACAATCACT 180  
25 AAACTTAAT CCAATGACA CCGTTTAT TTTCTTAAA CATTGTGAA GCTTAAGTT 240  
30 CTGAGATCA TGTGGCAAT GTATGGGCA GTAAATGAC AGAAGATG TTTATGACA 300  
35 ATTAAGGCT GTTTCACCT TTAAGACA GCTGGGCTT AGTATTCCT GGGGCCAGG 360  
40 TGCATTAAG TTTTACAAA ATTAATCAT ATGTCATG TTTGCATTT TTTTCTCTG 420  
45 TTGAATTTT GAAACGAG TTGACATC ATTAAGATA TTACTTTCT TCAATGCTT 480  
50 TTTGTTTAC TGCCTTAAA GCTTTCGAG AATATCTAG GCGACAGG CATTCACTT 540  
55 TCCATCTTAA TAGAATGAA ATTAAATTTG TATCTACTA TAAAGAAAT TGGGTCAAT 600  
60 GAAAAAAT CATTATACC GTCTTTAG TATATGTTA AAAATATAT TTATGTCT 660  
65 GAAATTTCA GAACTCTT GAGACACA GTTTCATAT AACTTTCT GACATATAT 720  
70 GCTGACAGG TTGCTCTCT TTTGGAAA GAAAGGTTG TGTAACTAG GCTTAACATC 780  
75 TTCAATATC CAATTTGTA TAGATTAAT AAATATTTA TTTTATGCT CAGTATTTA 840  
80 TTATTAAT TTTTATGTA TCCATCTC TTGCTCAT TAAAGAAAG GCAATCATTA 900  
85 GAGATTCAG GATATTTTG TTTAAATCT TCGAATTAAC ATGTTTAC AGTCCCTGC 960  
90 TTTTGGAAA AGTATTTCT CTATCACT TGTTTTACC TTTTGAACA TTGACAGAA 1020  
95 TTATCAATG GTTGTGAG ATACGACT GATGTGCTG TTTATCAT TTGCTTCAA 1080  
100 AGTCCCTAC TCAAGAGGC CTAACTAG TAGAAATTA AAGATTTCA AAACTTTCT 1140  
105 ATTCCTTCT TAACTTAC AGCAATGAG GATGTGATA GCAATGAGT GTATGATGA 1200  
110 GAAATTTGA CCAATTTGT TTTTGTGAG TTGTTGTAT TTTGAATTT AAATCATCT 1260  
115 TATTCCTTT AAGATTTT ATGTATGAT GTGAATATC TACCAATCT ATGCTCAAT 1320  
120 ATTAATGTA AGTCTCTT GACCTTAC AAGATTGAG ATGCTCACT GATGATGAT 1380  
125 ATTCTTTAG TAAATATG TAAATTAAC AATGATCTT TAAATGATG ATGCAATCT 1440  
130 GTATTAATG TGTGTGCT GCTCTAAT GAGGCATAT AAACAGTTT CATATATAT 1500  
135 TTTCAAGAT TGAATTCAC AACTGACT TTGAAATTT CTTCCATCC TGAATGACA 1560  
140 ATTAAGAGG CAAATATAT TCCCTCTTA TCTTGAAT TTAATCACT TTATGTTAAA 1620



449

450

AGTTGTGTAT AATTGTTAAA ATCTGTGAAA GAATAAAG TGGATTAAA TTAAAAAAA

1680

AAAAAAAA AAA

1693

5

(2) INFORMATION FOR SEQ ID NO: 222:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1196 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

15

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1196 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

60

ACCGTGGT CGACCCAGGC GTCCCGAGN TGGCTGTGT GCGAAGGAG AAGATTGT

120

AAACCCCGGA CGAGGTCTT GTTACCGGA GCGCGCTGCT GTCCGAGAC CCGCGGTGA

180

AGCCACGTC ATCATGCTG ACCAGGAGC AAAAACTTCA ACTGAGACT TGGGGATTA

240

GAAGCAAGT GAATATATTA AACTCAAGT CATTCAGAG GATAGAGTG AGATTACTT

300

CAAGTGAAT ATGACACAC ATCTCAAGAA ACTCAAGAA TCAATCTGTC AAGACAGGG

360

TCTTCAGATG AATTCACTCA GCTTCTCTT TCAAGGTGAG AGATTGCTG ATATCATAC

420

TCCAAAGAA CTGGAATGG AGCAAGACA TGTGATTGAA GTTATACAG AACAAAGGG

480

GGGTCAATCA ACAGTTTAGA TAATCTTTT ATTTTCTTC TTTTCCCTCA ATCTTTT

540

ATTTTAAAA ATAGTCTTTT TGTATGTGG TGTTCAAAG GGAATTCGAA ACTGCGACC

600

CATCTCTTGG AAACATCTGG TAATTGTAAT TCTAGTCTC ATTATTCATT ATTGTTGTT

660

TTCATGTGTC TGATTTTGG TCAATCAGCC TCAGTCCCTT TCAATTTACC CTCTCTTTT

720

TAAAAATAC GTGTGCACAG AGAGTGCACC TTTTTCAGGA CATTCGATTT TCAAGCTTGT

780

GTTGATTAAT AAGATCGACC AATCGAAGT TTTATATGA CTTTCCATT GCGCTGATG

840

TTCTAGCATG TGATTTACTT ACTCTGTGAC TGTGACTTTC AATGCGAGAT GGAATTTT

900

CAGAGCACTG AACTGTGAAA AATGACTT TCTTAACTT GAAGTACTT TTAATTTG

960

AGGCTCTGGA CCAAAAGAG AGCAATATCA GTTTCAGATC AAGATGAGC ATAGGTGAG

1020

AGTAATGACT AACTCCAAAG ATGCTTTCAC TGAAGAAAG GCATTTTAAG ATTTTAAAA

1080

ATCTTTGCA GAGATTCGA GAAAGTTCTT AATTTTCAAT AGCAATTAAT AAGCTATAC

1140

ATCCAGAAAT GATACACCA GACACTCTCT CTTTTCAT TTAATTTGAC TTTTGGCTT

1196

GGGATATGG TTTTAATGG ACATGCTG TACAGCTTC ATTAATATTA ACATA

60

TACTCAACG CTGAGATTC TGGAGCCAG TCTGCCATGC CAGGATGAC TGGACATGTT

CAATCTAGAA TCCCTTCACA CTACAGTCAT TTTCTTTCTT CTCTCTGACC CTTCGGCTCTT 1560  
GGAAATCTCG CTCTCTCAAC CCCAGAGCCT AAGAAATGCA GCGCTTTCTT AACATGTGTA 1620  
GAGATGATTC TTCTCTTGACC CTGGCCAGCT CCGGAGAGCTT GATGGCAATC CTGGAGAGGCT 1680  
TTAATCTCTT TTCTGTAGATT TGTGTGGGAA GGGAGAGGTGA TATGATTCCT ATTAAAAA 1740  
AAAAAGTATA TATGATATA TCTATATATA AATGAGCGCA GAAATTAATC T 1791

## 15 (2) INFORMATION FOR SEQ ID NO: 224:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2517 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

25 AACAATGGA ATCCAAAGAA TTGGGACAG CCGGACAGCA TTGTGAGCT TTCTCTGTGTC 60  
TTGGGAGCC TCAGAGGAGC TCGACTGTGTC CATCTGCGGG TACGAGATG CTGTGTGTGTA 120  
AAAAAGAGCA TCGCATTCAC GCAATATAGAA AGTAAAGACA CCGATGTGTC GAGGGGAGAGC 180  
CGATGAGATG CAAGCTTCAC ATGAAATGGA ATGTATTAC ACCTGAGCAG CCGCATCTGC 240  
TCCGAGCTGAG TGAAGAGCCA TGAATGAAA AGAGAGAGCA GCTGCCTGAC AGGATGAACT 300  
CTGCCTCTCT CTCGAAGCCC CCGTCCGAGG TGGAGCTTGA CAGCATCTG AAGGACACTT 360  
TGAAGTCTTC AAGGAGCTCT KTGACAGATC AGCCGACAGTA TTTCAAATTC AACCTTTGAG 420  
CAAGGAGATG AAGGAGCCAG AAGTGGGGGC AAGAGAGGCT GACTGTGTTT CCCCAGAGCA 480  
AAGCTTATGA CGAATGGACC ATTCAGATG AGAACCTTGA TTGTGGGAG GATTGCCAGG 540  
GATTAAGAGC TTCTCTCAGT GATGGAGCC GCTTTCTGT CTGTGTGTCT GCGCTGTGCT 600  
CTTCTCTCTA GATTAAAGTT TCCCTATGTA TGTTTCTTCA TCTCAATGCC AAGGTAAAGCT 660  
TGTGTTTTTA AGTGTGTGTC TCCCCAGGC TGAAGCCGAA GCTGATTTCT TATCTGAGAA 720  
TGTGATATG AATTCATGAG GTGAGTTTCT TGTGAGCCCA TGGATGCAAG CTTGAGGGGCT 780  
GTCTGAGAGA CCGTGCCTTT TCCAGGGGCC GAGGGGCTGC CTTTCCTTGG TGTATATTTA 840  
GCTTTTAAA CAATGAGAGG GATGAGAGAC CCGTGTGTCC TGAAGGAGGC CAGGTGTGACC 900  
TGAAGACTGT GCGGCTGTCT TGTCTGTCA GTGAGAGTGC CTGGGTGAGG AAGAGATGTC 960  
AAGGCTTTTG TCCCTCTCCC AGTGGCTTCA GAGCTTCACTA GTGGGAGAG CAGGATGAGG 1020  
CTGAGCCGCT GAGAAAGATC TATCTAAGCT CTGTGCTTGG AGTCCGCTGT GCTCTGACC 1080

CAAGAGATG TCTCAGATG TCACTTTCC CTTTCTCTTG AATTGTGTC AATGCCCCAC 1140  
CCAGACTCTC TTTCCTCTCT GATGTGCTTT GCTGAGAGG GCTGTGTGTC TGAAGCTTCC 1200  
CGATTCTCAC CTGGGCTGAG ACTTAAACAC AACCTGATT TGTGTAGCCG CCAGCTCTCT 1260  
TCTGTGTGAG CATTGAGAG GCTCAGCTCT CCATGTGCA GTGCTGTGAT TTGAGACTTA 1320  
TTGATATGA AAGATGATG TTGTCTCTG CTCTCAATT CTGGGCTGAG TTGTCTAGAG 1380  
GAGATGTGTC AAGATTCCT GAGGCAATAT ATTCAGATGC GACCAAGGG CACTGTGTGT 1440  
TCCGACTTAT GTAGATGAGC CCATCCATCC ATGACAGAGAG GATTATTTTC CTGCTTTGAC 1500  
AAGAGAGAG GATCTAGAGG AAGAGAGGAG CTCTACAGAG CAGGTGTTT CCCCAGAGTA 1560  
GCGGAGAGCA GTTGGAGACA AACTTCAGAG CCGAGGAGAT CCGTGAATGA CCGAGGCAAT 1620  
GTCTCTACAG AGTGTGCCCC TCGTGTGTC GAGTGAAGAG AATCCAGGCT GCGAGAGACTG 1680  
GAGCCAGTTG GAGAGACAG TTCTGGAGAG TCTGCAAAAT GATGACAGG TCGTGAAGAA 1740  
GCGAGATGCC GAAATATATC AAGAGTCCC AAGATTTGCT TGAAGCTTAC CCAATGAAAA 1800  
AAGCAGAGCA CTCAATGTTTC CAGGGCTGAG TCTGTACAGC AAGAGAGAC CAGATTTTGA 1860  
AACGAGCTTC AGTGTGAGG CTCTGAGGCT GCGCAGAGAG GAAATCTCA AAGAAAGGGC 1920  
CTGTGTGTGTC TCCACTTGA GTTCTTTAAA GATGTGTGCT TTTTATCTTC CTAAAGCTTT 1980  
CAAGTGGGTC CAGATTTCT GTTACAGCT GAGAGCAATT CTTGCCAGAC TTTCCTCTTC 2040  
CTGGCCAGAG AAGAGATCA GAAAGCAATG GAGACTGTTT TTTCAGATAC TGGAGAGCCG 2100  
GAGCTGACTG CTTCAGATCT GCTTAAAGGTA GGGATGTGTA AATATCTCCC TGCATGGCTT 2160  
TATCTCTCTT CTCAATCCAA AAGCAGATATC TTCTGTGTTT CAGAGATTT CATTGATGTC 2220  
AAGTGCAGCC AGTGTGCCAT CTGAGACTG TGTATAGCT GACCATCTG TACATTTGAGG 2280  
GAGACTGTTT GCTCTCTCCA CTCTATAGC AGTCAATCTG GAGAGCCGG AAGAGAAAGT 2340  
GATGGGCTAG TCCCTGTGTC TCTCTCACTT CCGCATGCTC TATGTTAACC ATCTGTGTCT 2400  
CCATGTGCA AAGAGAGAA GGGGCAATTA GAGATGAGG GTGATTTGCT ATTACTATTC 2460  
CAATCTGAA TTAACATTTG TTAATCTTAA AAAAAAAAA AAAAAACTGAG AAGGGGG 2517

## (2) INFORMATION FOR SEQ ID NO: 225:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2474 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

5 TTTGTACTTA TCGAGGATG ATTCTAATGA CAGAGTCTTT CAACACTTTG CACATGATGT 60  
ATCAGCAAGC TACAGCTTGC CATGTGACTG GAGATTAGT AGAATCTCTG TCATATTTC 120  
TTTGGTTT GAAGTCTACA CCGCTTATC TTGAGAGAA AGATGTGAAA CAAGCATTTA 180  
TCCATGTGGA CGAGCGAATT GAAATTGGCC ATAACTGTT AACTCTTCTT AATTCCTATA 240  
10 GTCTCCAGA ACTTAGAAT GCGTGTATAG ATGTGCTCAA GGAATCTGTA CTTTGTAGTC 300  
CCATGATT TTTTCACTACT CTGTTGCTT TTCTACACA CAACCATGT ACTTACCATC 360  
ACAGTAAAT ACCAATGTCT CTGGAATTT ATTTGCTTTG TCRGAAAT ATCAGCTAA 420  
TAGAGGGA AGCAATATT CCGCTGCG CCGCTGACT CATATGTGC CTCTTGCCA 480  
CAATGTGGA AACCAGTAAG GGCAGAGATG ACCTTTATGA TGTATGCTG CTAGACTACT 540  
20 TCTTTCTTA TCATGATTC ATCATCTAT TATGCGAGT TGCATGAC TGTGAAAT 600  
TTACTGAAC ATTAGTTAG CTGATGTGC TATGTGCTTA TGAAGTTTG CCACTGTC 660  
TTCACTGTT CCCCACCTT TCGATGAGC TATGCGAGC TCAATCTGCT ATGTGAAA 720  
ACTGATCAA GCTTTTGTGT GAGATGCTG TTTTGGAGA ATATATATA TGTATCTTAA 780  
TGAATGAG AACTTTTTA AACAGACA TTGTCTACAC GTTCATGACA CATTTCTTC 840  
30 TAAAGTTCA AATCAGATG TTTTCTGAG CAATCTGTGC CATTTGATC AGCACTCTA 900  
TTACAACTT GATAGCGAG TATCAGAAC TACAGTCTGA TTTTGCAC CAGATGAAA 960  
TTTCMAAGC AGTGTCTCT TTAATGGGG ACTGAGGGC ACTGCTTTG CTCTGTGAG 1020  
TACACTCC CAACAGTTA ACCGAGTC TAAITCCAC TCTGCAAG CTTTATAGA 1080  
AATGAGAG TTGTCTGGA CAGAGAACT CACTCCAGA GCAGAGGC AAAGAGAA 1140  
40 AACTAAGA TATGAGGA CCACTTCCA TTAAGGGG CCGTGTAGC AGTCATGAG 1200  
ACCACTGT AGACAGTGC ATCATGACA TGAAGAGGA AACGAGGAG GTCTGAGCC 1260  
CAAGGACAC TTCTGACAT GAGAGCAGAG ACTGCTCAAT TATTATCCA GGAATGAGC 1320  
45 AAGATCTCC TTGCTGAA ATTAGTTCTG TTAAGAAATA CCGAATGGA GTTCAATCTT 1380  
CGTTTTCGA AGCATGTCA ATATAGGT ATCAGATGC CAGAGCTGC AGAAGAACG TCCAAATG 1440  
50 GTAGATATGA CGATTGTAA GATTATAG ACCTGCTCTG TTCCAGGAT TCTACCTTAG 1500  
CCGAGGAGA ATCTGATTC CTTTCACTT CTATCTCTG ACTTCTGCT GACTTAGCTG 1560  
ACTTGAGAG CTGTATGCG CAGGTTTG CCGTCAGGA CCGTATGCT GCTTTATCTC 1620  
55 TCAATGTG CCAITTCAGA GCACTTTTA GTCATATCA GCACATGAC ATTTTATATA 1680  
CCCTGTATG GACCATGAA TCTACATGC ATGTGCTCAC AAGGATATC TGGCAAGGA 1740  
60 AACCAAGCTC CTTCTGACA TTAGGTAG CATGTCTACT TTTAATGCC TCACCCCAA 1800

CCCCATGCT GTTGTATAA GTTTGCTTA TTTGTTTTG TCTTTCAGT TCTCCAGTGC 1860  
5 TCTCTCTTG AATGCCAGA TAGATTTATA GCGTTAATTC TTGTCAGGC AGAATCTCAG 1920  
ATGAAAAA CTTGATCTT CAGTACTCT CTTAAAGGGC ATTCAGATTA TGGATATGTT 1980  
TTATGTAAIT AAGAGTTCAC TTTAGTGGCT TTTCAVTTAAT ATGCTGTCT GGGAGAGACA 2040  
10 GGGTGGCTA GCGCTGTACA ATGTAAATTA AACTTACAGC ATTTTACTG TGTATGATAT 2100  
GGTGTCTCT GTGCGAGTTT TGTACCTTAT AGAGCGAGAT TGCCTCCCAT CCGTGTGCTT 2160  
CTTATATCA AATTAAAGTT TACTTGTATA CCGACACACC ACAGAAATTT TGAATCTGTA 2220  
15 AAGANTCTC TTTAGCTGTG GCGTGGCACT ATATAAATGG TCGTTTATTT AACAGATAC 2280  
CTGTGGAGA AATAAGCAC ACTTGATGTA AAAATTAATG TTTTATTTTT ATTGACATGA 2340  
20 CTGATGATT GCTATCTGT GCACTAAAT AAACGTGATG TGAATGACTTA AAAAAAATA 2400  
AAAAAATA AAAAAAATA AAAA 2424

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 226:  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

5 ATATAGAGAG GATATCTGT TTAATCTCTG TTCTTCTGA TGCATCTACA AGCGGTAC 60  
TAGGTGACA GAAGCAAG ATCTTATTA AAAGAGTCT TACAGCAACC CAAGTCTCA 120  
TCTTCCATA GTAAAGTGA CCGCGCTTG AGGTAACTA CAGGCACAC CACTTCCGG 180  
TTTCTCTTC CCGCTGTCC AAGATGGCG ATGAGGCCAC GCGAGTGT TGTCTGTAGA 240  
45 TCCCGTCT GAGACTAAC CCGGAGCCC GAGATCTGTA GTTGTGGTG CAGGACTGA 300  
AGGAGATA TCACTGCTT ATCGGTATG TGGAGAACAA CAAGAATGCT GACAACGAT 360  
50 GGTTCGACT GGAATCTAAC AAGAGGAA CTCGCTGTT TGGAAATGC TGTATATCC 420  
ATGACCTCT GAAATATGAG TTTGACATCG AGTTTGACAT TCTATATCA TATCTCTATA 480  
CTGCCCCA AATTGCAAT CCTGAGCTG ATCGAAGAC AGCAAGATG TACAGGGTG 540  
55 GCAAAATAG CCGTACGAT CATTTCAAC CTTTGTGGG CAGGAATGT CCCAAATTTG 600  
GACTAGCTCA TCTATGCT CTGCGGTGG GTCCATGCT GGCATGTGAA ATCTCTGATC 660  
TGATTCGAA GCGGCTCAT CAACCAAG AGAATGCAA CCAATGAGA ATCAAGCCAC 720

780  
840  
900  
960  
1020  
1080

TAAGCCAGGG CAGAGGAGC TTGATTAAGC TACGATACTA TTTTCTCTTG CATACAGCTT  
AACTGATCTA ACTGCTTCCC CGGACAGCTT CCAGCTTAG TTGTATCTAA GTACGTGAGC  
TAGGCACTGC TGGGAAAGAA ACAGACAGC AGCAAGAGT ACTGCTACTT AGTTTCTTAG  
GCTGACAGG GAGGAGAAAG ACTGCGCTTT GAGATATCTA GAGGTAAATT ATATCCGCCC  
CGAGCTGAG CAGATGCGA TTCTGAGGC AGCGGGGTAA CTGAAAGTGA GTACATTAAG  
TCTTTCTGCT TTCTGAGAT AACCATCTAA TAAAGCTGC TTCTCTGAG TAAAAAAAG

15 (2) INFORMATION FOR SEQ ID NO: 227:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1316 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080

TTGCATTCAC AATTACTGAG AGGAGGCGAG GGGCAGTTGC ATGCTGAGGG TGGCTGCTAG  
GCTGCGCAGC TGTCTGAGGT TTGAAAGATG CGGTACAGCT GCTTCAGCTG AGCAAGATG  
TTATCCCTGA TGTCTGAGGT TGAATCTGC AGCGAGAGC TGCAGTATC AAAGATCTGT  
GTAAATTCAC CAGAAATCAT CTCTAGATC ATCCAGCGA CTACGTGAAT GACTGAGAGC  
AAATGAGATT TCTTTTCCAA TGTTTTCCAA TCTGGCGATT CTTCGCCAAA GCATTAATAG  
ATTCGAAGG GTGCTGCTT CTCTATCTGT CCTTTCTGAT GGGCAATGAG ATCCGTAAAG  
AATGTTTCCA GACAAATAG CTTCAGCTTC TTTGTCTCT CATTCAGATT GGAAGCAACA  
AGTATAGGG CAGATGCCC AGACAGTAC AGCTTGCAGT GGCAGAGCTT GATGCGATTA  
ATGGATTCAC CGGTACATC CAGATCAGT CCTCTGTCCA TGAATCTGAG CAGCTTCTCA  
GTGACAGCT TCTGCTTCT ATTTGATATA TGTCTAGAG CTGGAAATT GACTGCTGCC  
AGATGACGG GACGAAAGAG CTCTCTCTGG TCAAGCATGG GAGCCAGGTC CCGATTAGAG  
AGTGGCAGC CCTGAGGGTT GCTCAGGATC ATGATCTGTC CGTATCTCT TCCAGCGATA  
CTGAAATCTG ATGTCAGGT CAGTCTTGG GAGAGAGTC ATCCAGATTT CTGAGAGCT  
ATAGAGAGTC TGTATAGATT CTGAGAGAG TTGCATCTCC AGGGTTGAG TTTTGGGACA  
CAGTGCCTCC GAGTGCAGT TCCCGACAT GCATTTGCC AGACTGCTG GCGCAGTAGG  
AGAGCATTC ATGTTCAAGG AGGGAGAGT GTCTGAGAG GAAACATGGT GCTGCGACATG  
ATCCAGTCA TCTTCTCAT CTCTTCATC CAGATCATTA TCTCTCAT CCGAGGAGAGC  
AGACCTCTG GATCTGAGT TAAATATCA GCGCTGAGGC TGAAGGAGAT CACACACTTC

1140  
1200  
1260  
1320  
1380

ATATATCTC ACTGGATCTA TGGGAGATC CCTGTGAGC ATCCATGATC CAGATGTA  
TTCTCTGCT TTAATAGAG CAGAGCGAG CTGGCGCTTC CATTTAGCT GCTTCAGGCTC  
ATCCAGCTCT TCTGTAGCT TCCCTCTCTC TACAGCCGAG GCGTTAAAA TGGTATTTTC  
CTCTCTTGT TGAAGCTAT GCGGGGTGAG ATGTTTCTAG GAAATCTGGA AGCGTTTGA  
GTCCCTGCTT AGCCAG

15 (2) INFORMATION FOR SEQ ID NO: 228:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2043 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080

TGAGCTGATC CTTCTCTGT GTCTGAGGG AGCTGCTGAG GAGCTTCTCC TGGAGAGCAT  
GAGCTGACAG CCGACCCACA CTTCAGATG AGAGCTGAG CTGCCCCCGG CAATATGCTT  
CCGATGCTCT TGAAGCTGC ACTTGCACA TGTCTAGCC CAGCAGATC CGATGAGCCC  
CTGACCTCCC CTTCCTGAG CTCTTCCCA AGAGATCTG GTACAGCTCT GCTGTGCAAT  
CAGAGGCCCA GGTCTGACA GCTCCGSGGG AAGCGTGTCT GCGTGTCTCT CGATGAGCT  
CGAGTTCAG TCTGAGAGC GTGTGATGA AAGTCTGAG CAGCAGCTAT GCGCTCTGAG  
GAGGCTGAG TCTTGAAGG CTGATTTTTC TGCAGAGAG TTGAGAGAG CAATGCTTT  
CTGAGCTGAG AGGATCAGCT GCTTCAGCTG GGGCGATGC AGAGAGGAG GTGGGCTGTG  
CTGAGCTGAG TCCGAGATG GCTTCTTAC CAGAAAGCC TCAAGCTTCC TCTGAGACA  
TCCCGCTTTC TGGCAGAGG GAGAGGCTC CTTTAAGGG TGTCTTCTCC CAGTGGGAG  
CAGTCTGAGC CTGCCCCCTA CTAAAGCTC TGTCTGAGC ACTTTTCCCC AAGTCTTGT  
AACTGCTGAG AAGGTGAGTT TGGATGCGA GCGAGTCTCT GAGCAAGCTC TCCGCCCCCT  
TTTAAATTC ACTCATTTTG TAAAGACCA GCGAGCTGAT GTTACTTGA CCGTGAAGT  
TTTTCATTT TTTCTTTCG TCTTCTCTCT TGAATGAGG GTTCAATATT GCGTCTCTCC  
CCTGATGAG GAGATGCTCT CTTTCTGAG CCACTGAGCG GCTGTGCTCA GCAAGCTGAG  
TGCAGAGCTG GGGGAGCGAG ATGGGGGCTT CTCTCTCTG GAGGGTTCGA GATGCGCTCC  
CGAGCTGAGG AGGCTGCTTT CCGTACCTCC CCAATGTCCC CCGCTGATGA GATGTGAGCT  
TCTTGTCTCT GGAATCTCT GCAATGAGA AGAGAGATTT TCCAGATTT GTTGTGTGTT

TTTTTCTAC CTAACTCTTA GAAATATGAT GTTAGAGAGT GCTTGGGAG GGGGAGACAG 1140  
GTGTTTGCTC GGGCTGAGAG AGGCTCTGCT CAGCCCTGAG AGTCCCTTCC TGCCCAACCG 1200  
5 ATACTGGCAC TTATAAAGG AAGCTGACCG CACAGTGTCC AGACGATTC GCGCCACAGA 1260  
GATGGGAGT TCTGTCTTGC CTTTCTGTGT CTGCTGTACC TCACCAGCC TAGAGGGAG 1320  
GTGCATTCAG GTTAGATTTC CTTCTATTC AAGTTCTGG GCTTTGGGT GGAACACAGC 1380  
CAGCTTTGGC GCTGTTGGGG AGACTCTCC AGACCAGGAA CCCCAGAGG AGACAGAGCC 1440  
TGCCATATC TCCCAAGCCA GGGCTGGGC CAGGCTGATT GCACTGAGAA TTTCGCCACA 1500  
15 ACCAATTGA TCTTGCTGG AACGAGGC CAGAAAGCT GGGCTTGTCC CAACTGTGGA 1560  
GGCTGTCTT CAGCCCTCTT GTCCCTTGA GCTCAGTGA TTCCACCAG GTGCCACAG 1620  
CTCTGGACT TCAAATTTCTA TATATTGAGA GAGTTGGAGA GTATATCAGA GATATTTTTC 1680  
20 GAAAGAGTT GGTCTATGCA ATGTCAGTTT GGAATCTTCT TGAAGTTTAT ATGTTTATAT 1740  
TAGGAGATT AAGAATATA AAGCTTACA ATATCTTTAG GTTTTATTTT TTTCCTGTTC 1800  
ACCCACAAA CTGACCAANT GGCATGTCTA TCAGATGGA GGTGTCTCAT GTTCTCTCT 1860  
25 GTCTTAGGG AGTGATTAAG GAGATGGGSG AAGGGGTGTT TTTTCTTTTG ACTGCCCTCC 1920  
TTTCTAACAG AATGTTGCCA CCACTCTTG AGTGGCTGT GTTTTGTCT CTGTCCACG 1980  
30 TTCTTTTGA GAAATATACA TTGTTAGGG AACTCAGCT AGTGTACCG TCTTGTTTC 2040  
GGG 2043

## (2) INFORMATION FOR SEQ ID NO: 229:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 540 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

TAAAGAAG CCGAGAAATC TGCGCTTCC TCTAGAGATC GATGGCTAG AGGAGAGCT 60  
50 GTCCATGTT CCGAGAGACC TGGAGCGGCT GAATCTCAGA CTCACAGCC GGGAGCTGAG 120  
CCAGAGGCC AGGAGTCCC TGGAGAGGA GAAAGACAG CTATGACCA AAGCTCCAA 180  
CTAGAGAGG GAATGAAGT TTCTTGCCA AGAGAACCG AGAATATCC TCTCTCTCT 240  
55 GGCATTTT ATCTCTCTGA GCTGTCTTA TGCTTACTGG ACCATGAG CTTGCACTT 300  
CCGACAAAC AGCAGAGGCT TCATCTGGC CCGTTGGTCA GCATCAGCA GGCATTTCAA 360  
60 GCTTCATAG GACCAAGTG CTGGGCTGTT CCGCTTCCAA CTTAGTGTTC AAGCATGCT 420

TCCTGGGGGC CAGGCGTTG CTTCCCTTGC CTCTGGGGG GTTCCGGGTC TCCAGAGCA 480  
CATGTGCTGT GTCCCTTCTT TACCCCAAGG GAGAGGCAAT AAGACACAA AGCTGGAAAT 540

## (2) INFORMATION FOR SEQ ID NO: 230:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 448 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

AATGTGAAA TATTAGATA TTGTACTAT TTGACCCAC TCAAATCTC CATGGAAAA 60  
20 TACTGTGGA TACCAAGCT ATGTGTGAA ATATCAGAT GCAGTATAC AGCTGTGTA 120  
GACTCTAGTA CCAGTTGGGC ATCAAGCCA CAGCTAAAA TTGAAACAA AGATCTGGAC 180  
25 AACAAACAG CCAAGGTGG GGTTCAGAA GCTCTAGCT GTACTAGCT GTAGAAATCT 240  
ATGCAAGCT GCCAGTGTG GTGTGATAT CAGGAAAAA CTCAGAGAG CCCAGTCTT 300  
30 CAGCTCTGT TGAACATGAG CTCTGTGTA GCAAGAGTG AAGGCTAAG CAGATTTAG 360  
CTCTGAAGC ATTCACAAAC ATACACAAA ATCTGTCAA GCATTAGGA AATCTTGTTA 420  
CTGCTAAGTG TTCTGAGCC AGAAGCA 448

## (2) INFORMATION FOR SEQ ID NO: 231:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 407 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

GTATGCTGCC CCAAAACAAT ATGTGTGGCT GCCTTWAAC TGACTTCTCC AACATGTAGC 60  
50 CCGACAGGA GCGCTTAGA CTRAGGGAGG GCGTGTGAC CAGGTGTG TGCGGCTGCA 120  
TGAACTACC AGAGAGACAG ACATTTCTGA ACTCACCTG GCGATTCAG TCGATCTGCC 180  
240 TATGCTTGG TCCAGCCAG ACCTGTGAGA TTTCTCTCAT GAGGATGAC TTGTGCTTCT 240  
55 GGAAGTTTGT CTGCAAGCTT ATAGTACTC CCACACAC CAGCAATAGA GTTAGCTACC 300  
TOTGGCCTTG GATCTAGCC AGCATGGCTG GAGAGAGGAG CAGCTGGCA TGTACCTTAA 360  
60 ATGCTCTTAC CAGGAGGGA CTCGAGAT GAAGCAAGT AGCGACT 407

## 5 (2) INFORMATION FOR SEQ ID NO: 232:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 830 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

15 GTATTGATT TCAAGCTGCT AATTGGGCTC ATTTCAGATT CATTCCCTGA TGTAGACATT 60  
AAAAAAAA CTGATAGACA TTCTTTCAG GATTAAGTATT AACAGACGA TGTAAAGCTT 120  
ATTAAATGAT CAGACATGCA GAGACGTTT AAGGTGCCA CGAGAGAGA TCAAGACATA 180  
GAGACAGTG TACCTCGGTC TGCTCAGAG ACCGAGACA TGTATCGATT TTTGTGTG 240  
GTTTATTATT TTTCTGTATA AAAATTGTGA AAGTTTGT TTACGTAGAT GATTATTATA 300  
TACCTCGAG TGCTTTGGA CTATTAAGAT GTCACTAGCT AACAGACATA CATTATGTTT 360  
TTTTTGTGTT TGTATTACAC TCAGTATATA TCAAGAGAG TTACGCAACC ACTAGACATT 420  
TGAATTCCTC TTTTATTATG TCTTTAGAG ATATGAGATG TCCATATACA GAGACAAAC 480  
ACGACAAAA ACATTGCTGA ATATATAGCT GATAGAGATG AAGCACTGCG TTAACTTGTG 540  
GTACCAATAA TTTAGTGTGT ATATATGAT ATATATGAC ACACACAC ATATATATTC 600  
AACAAATATA GCAAAATATA ACATGCAATT CAGCAATTGT CTTTCCCTGT TACGATTTTA 660  
ATACGACAC TGTATAGATA GTTTAGGTGA TGTAAATTC AATTATATGT 720  
AAACGAGATT AACAGACGA AAGAACTGT CATTATGAT GAGTCATGCC TTTCTATTAT 780  
AATTAAGTGG CTTCGGTTAT CGATCAATAG CAGCAATTATA CTGTATCTG 830

## 45 (2) INFORMATION FOR SEQ ID NO: 233:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

55 CGAGAGAAA GACCAATCTA GAATATGGA CTCTATACAC TTCTAGTATT TCAAGTTCTT 60  
AAGCAAAATG AACTTGACCC TACAGCTAGG GCAATAGCAA TTTTCTTTAT GTAGCAATG 120  
CTACGGAAC AAAAGAGTG AAGAGAGCC TTTTATTATA CTTAATGAC ATATATGAC 180

TTTTAGACA AAGATGGCAG AATAGGCTT CATTCTAC CCAGAAAAA ATCCAGATT 240  
GCTTCTAAA ATGATATCAG TTTCTAAAGT GAAAGATGCA AATATATGC TGTAGATGAC 300  
TCTGAATG GAGAGGAGG TACTCTGTG TACAGAAAC TGTATTGTGA TATATGTCAG 360  
GCTGTGATT GTAGCTATCA GCAATCTGT GCAATATAC TTTTCTCAT CATTGACTGT 420

10 GCAAAATGA TACTTTAAA GCAATCTTT CTATAGAC AGGTCCTCT ACTGAAACTT 480  
AATTGACAA AAGGTGTCAT ATGCTTCTCT AACCTGATT GTATTAGAT TCAAGAGCC 540

15 TACATTCTC CATTAGGTT TGTATGCTCA GTATCTTTC AATGCGCAG CAGAGCTTAC 600  
CTTTCTGAT GAAACATACC ATTATTGTA TTTGACAGT ATAGACATGC ACTTGTGAG 660  
TCCGATTTA AAAATGACA ACTGCTTAT CGAAGATGC TGAATAATAC TGTCTATGCC 720

20 AAGTTCTTA AACTATAAA GCAATTTTG CTYTGTTG TTATATCAG GATAGGCCA 780  
GCAATGCGA TTACCTGAG CTTTAAATC AATATGATG CTGTAAAT GACGACTGTC 840  
TGCAATGCA GAGCTCGAG AAGCAATTC CAGACTGTG ATTGCCAGA ACACATAGTC 900

25 CCGAGCTTC TAAATGAG CAAATCTAAA AG 932

## 30 (2) INFORMATION FOR SEQ ID NO: 234:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2786 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## 35 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

40 TTACAGAGGT GAGCTGTATA AACAGACAC ATCTCTATC CCTCTTCTT TTATTCCTCC 60  
CTGGTTTCA GAAAGAGAG ATATATGAG AGCACTGCC CCTTCTTGA TCCGAGATC 120  
TCACTGCTCC TCCCAAGCTT CCAATAGCT CTCAATGTG CTCACTGCT TGGAGAGAG 180  
CTCCGATAG GAGAGAGCT GCGCTCTACA GTCTCTTGA CTGTAGACA GAGCTCTGTA 240  
TCAAGAGAC GATGAGAAA GTTCCAGCT AATGCGACA ATTGCACTT TGAATATG 300  
TGTTTTGTG TTGTGAGAC TGAATCTCT TATTATTATA CAGAGATCT GATTTTTTT 360  
TTTGTGATGC TTTGTGTCTA TATTTTGAG GCTGTGAGA GAGAGTTAG ATTATTGTA 420  
GATGAGATCC GTTCAATAC AGTATCTTG AAGCGAGAC ATAGGTTGA AGAGCGACA 480  
CGAGCTCTG AATATATAC TGTCAATGAG CTTTAAAGA AAGCTGTCC TCAAGCTTAA 540  
CAAAATCACT ACAATAGCTT AGTCTTTTT TGAAGCTT TTACAGAG AATGTAGCT 600

WO 98/39448	PCT/US98/04493	461	462	PCT/US98/04493
5	660	720	780	840
10	900	960	1020	1080
15	1140	1200	1260	1320
20	1380	1440	1500	1560
25	1620	1680	1740	1800
30	1860	1920	1980	2040
35	2100	2160	2220	2280
40	2340	2400		
45				
50				
55				
60				
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 458 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear				
(2) INFORMATION FOR SEQ ID NO: 235: (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 235:				
5	660	720	780	840
10	900	960	1020	1080
15	1140	1200	1260	1320
20	1380	1440	1500	1560
25	1620	1680	1740	1800
30	1860	1920	1980	2040
35	2100	2160	2220	2280
40	2340	2400		
45				
50				
55				
60				
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 591 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear				
(2) INFORMATION FOR SEQ ID NO: 236: (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 236:				

180 TTTAGCTCT AGATTGCTA TGAAGGCTA ACTGAGTCA AATTGACTT GATCTCTCTG  
240 AATTGACTT GAGCTGACA ATGAGAGAG AAGAGAAAA TGTGATGCA TGTCTCTCC  
300 GAAAGTATC ATGAGCTTTTG GATTGCTTT GATTATTTT TCTTTTTTC TTCTCTCTCC  
360 TTTATGAGCC TTGAGGACAT TGGAAATAC GAGCAACTC TCCAGATCA ATGTACTCC  
420 ATGAGACTG CTGCTCTTG TGGTGTATC TAAATTTTG GATGAGAAA GAAATCTTT  
480 TGAATAAAA TAAATACAA AACATAAAA GTTTATTAG CAGAGTGA GCTTGAAAAG  
540 TTTTGTCAA ATGAGAGAG AGAATATCT TTTTAAAGG TACATATGT GAACTATAT  
591 GTACAGTGA ATAAATTTGA AAGTGTAT TTTCCAACT GTTGGAAT T

## 20 (2) INFORMATION FOR SEQ ID NO: 237:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1286 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

60 TCTTTTAG GTACAGCAG GAGAGCTAG AACTGAGAG AAGAAACTG CCTCTCTATC  
120 TACAAAGCT GATTACTT CTCTCTCTTC TTTGTGAAAG ACTGGGCTTC CAGCAAGAG  
180 GAAATACTG GGGGCAATG ATGTATGAG TCGAGCTATA ACTATGACC GATGAGAAAG  
240 CAGGAGAGG GCAATGAGA ACAGCAAGT ACAGTCTCT TCTGAAGAT CTGTACTGA  
300 AATGAGAAC AATTAGCA AACAGCTCC GTTTTCTCT CAGAGACTC CTCCAGTCA  
360 CTTTCACTT CTTGCAATTC TTCCAGCTG TCCAGCTGC ACAGTCTTC GACTCTGAT  
420 TTCCAGACG GATTTTCTC CTCCAGCAG CAGTCCACT CAACTCTTA TACCAAGAT  
480 AAGAACTGA CATTCTCTG GTTATGATG TGGTTTGA CTTGCAATTC CATTGAGAA  
540 TCGATGAG AGCATGAG ATACAGGGA TATGAGAAA GAGTTATGA GCTTCAGAA  
600 GCAATGAG AATAGGAGA ACAGCTAGA GAAAGAGAC ACAGGAGAA AGAGAAACC  
660 AGCATATGT CTTCTGAGG TAAATGAGA CTTCCAGTG AATGAGAA AGAGATATG  
720 GAGAGAGAC ACAGAGCAA AATATCTAA AGAGAGAGG AAGAGAAAG ACGGCACT  
780 GAGCTGCCC CTGAAAGGA GAGAGCCAA CTAAGACTG CAGATAGAG ATGCTTTTG  
840 CATTGTGT AATATGATC CAGAGTGA TACTAATAAT CTTGTATTT TTCTGAGAA  
900 TTTTAAAG AATTACTTA AATCTGTTC TGTTTTAG TATGAAAGT TAACTTTTT  
960 TCCAAATTA AAGAGTAT TTTGATGTT AAGTTAAAA TCTTGTCTT GACTATTTT

1020 AAAAAAAA AGAGCAAGT GACTTATAT CAGAGAAAG ATGAGATG ACCTACTTA  
1080 CAGAGATCT AAGAGCTGT GTTACTGTG TACATGCA GATTATGCA GAAAGTCA  
1140 AAGGTTCAC TTGGCCACA GTTTTGTGT TATCAACA CCACTCTTT AAGAGCTGC  
1200 ATCAAGAG GCAAGCAGG GAGGCTCTT AAGCTTTGA GATTAAAC TACCTTTAT  
1260 CAACTACTC TGTGAGACT CTGAGCTTG TATTTTGA GGAATCTTT ACTTTTTTT  
1286 GATTTCAC ACTTTTTG TTGGGC

## 15 (2) INFORMATION FOR SEQ ID NO: 238:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 734 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

60 ATGCAAGCC AGAAGACAA GAGAAAGAT GCGAAAGCC AAGGCTGAG CCGAGCAAC  
120 CTGTGCGCA AGCTGATTC CTCCGATGA GCGGAGAGT GCTTGAGAG GCGGCGCGG  
180 AACATCGGC CTGAGAGAC CTTCGTGAC CAGAGAGCT TGTCAAGCC CCGAGACTG  
240 GAGAGCTT GCAAGGCTT CTGAGAGAC GTGAGTACT ACAGAGCAA CTATGTCTC  
300 GTGTCTGAG GACTGATCT GTACTGTGT GTAGCTCC CTAATGCTT GGTGCTCTG  
360 GCTGTCTTT TCGGCGCTG TTAAGTCTT CTAATGCTT ACCTGAGAT CAGAGCTGT  
420 CCGTTGAG GAAAGTGA GCGAGGCAA TCAATGCTT TGGTGAAG GATCTCTTC  
480 CCTTCTCT GACTGAGTG TCGGAGCTG GCGTCTCTT GGTGCTGAG AACCACTTG  
540 GTGTGATG GCTCCAGGC TCCCTTCAC CAGATTGAG CTGTGAGAG GAGAGAGCTG  
600 CAGATGAGC CCGTGTGAG TGTCTCTG GACTGAGG CTTCCGAGG CAGCTGCCC  
660 AACCTGCCC ATGCTGTGC TCGAGGCTT GCTGCTGAG CCGAGAGGC CTTCCAGTA  
720 CAGGCGGAG GAGAGTCC GCGTTGAAA AATAGCTT TATGAGTGC ATTCAAAAA  
734 AAAAAAAA AAAA

55 (2) INFORMATION FOR SEQ ID NO: 239:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 809 base pairs  
(B) TYPE: nucleic acid

60



(C) STRANDEINNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

5 ARGGGCTGCA GGGCATCAGC GGGCTGTCTC CGGGCGTTTC GGGCTCTCAT GTGGGGGTGC 480  
TGTGTACAT CATCTCTCTG AACTACATAG ACCGGTTTAC CGTGGCTGGG GTCTTTCGG  
600  
5 ACATGAGCA GTTCTTCAAC ATCGGGGACA GTAGCTCTGG GCTCATTCAG ACCGTGTTC  
TCTTCAGTTA CATGTGTGTT GCACTGTGTT TTGGCTACCT GGTGAGACAG TACATTCGGA  
720  
10 AATATCTCAT GTGGGGGGG AATGCTTCTT GGTGCTTCTT GAACTGTGGG TCAATCTTCA  
TCCCGGAGA GCAATTCCTG CTGCTCTCTC TCAACCGGGG CTGTGTGGGG GTCGGGGAG  
780  
840  
15 CCGATTCCTC AGCATCTTCT ACTTTCCTAT TCCGTTGGG AATGTTCTGG GCTACATTCG  
AGGCTCAAA GTGAGGATA TGGCTTGAGA CTGGGACTGG GCTCTGAGGG TGAACCGGG  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1500  
1560  
1620  
1680  
1740  
1800  
1860  
1920  
1980  
2040  
2100  
2160  
2201

(C) STRANDEINNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

5 CGGGGTCTTC AGGTATACCG GTGTGTACCA GAGCTCTTAC CCTCAGCAG GCARACATCG 60  
CAGCGCAGAA GGAACACAGG AAGATATCGG AGCGGAGAG GCTGAGCGGC AGACCTCTG  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
809

(2) INFORMATION FOR SEQ ID NO: 240:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2201 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEINNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

5 TCGACCCAGG GTTCGGGCA CATGGCGCT GCGGTGTGTC AGCGCGGGG CTGAGCGACA 60  
GCAATGTCAG GGGGTCTCTA CCGCGGTGA GGGTGGCTT CCGCTGAGA TGTGGCTTC  
120  
180  
240  
300  
360  
420

## (2) INFORMATION FOR SEQ ID NO: 241:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

60 GTCTTCCCGG AACTGAGACA GTTCTGAC ATCCGAGACA GTAGCTTGG GCTGATCCAG  
120 AACCTGTACA TCTGACGTGA CATGCTGTGG GCACTGTGTG TTGGTACCT GGTGACACAG  
180 TACAAATGGA AGATATCAT GTGCGCGGAG ACTGCTCTCT GTTCCCTGCT GACACTGGAG  
240 TCACTGTACA TCCCGGAGAA GCAATTTGAG CTGCTCTCC TGAACCGGAG CATTGATGAG  
300 GTGCGGAGAG CCAATTAATC CACAAATGAG CCGACTTACA TTGCGACACT GTTGTGTGCG  
360 GACAGAGGGA SGTGATGCTC AGCAATCTCT ACTTTGCGAT TCGGATGAGG AGTGTGATGAG  
420 GCTTACATTC AGGCTTCCAA GTGAGAGATA TGGCTGAGAA CTGGACATG GCTCTGAGAGG  
480 TTGACAGGAG TCTAGAGATG GTGCGCTTTC TGTCTGCTTT CTTGTAGTGG CCGAGAGCGC  
540 CAAGCGAGAC CTGAGAGGCG CACTGAAATT TCCGACCTCT GAAACCGAGC TGTGTGTGAG  
600 CAGATTTGAG GCTCTGTGCA AAGAAATCTA GTTTGTCTCT GTTTCCTCTG GCTTTCACATG  
660 CTGTGACCTTT TGTGAGAGGCG TCCCTGAGCTC TGTGCGCTCC GCAATTCCTG CTGCGCTTCC  
720 GAGTGTGCTT TTGGGAGAGC CGAGCTTCCG TTCCCGAGAA CTCTGCTGCT TCCCTGTGACA  
780 GTCTCATCTT TGGACTCATC ACCTGCTTGA CCGAGATCTT GAGTGTGAGG CTGAGGTGTGAG  
840 AAGTACAGGAG CCGATTTCCG CACTTCAAGC CCGCGAGCTGAA TCCCTGTAGTC TTGTGACATG  
900 GCTCTCTAGG CTCTGCAACC TTCTCTTTCG TTGCTCTTCC CTGCGCTTCC GGTGACATGAG  
960 TTGGACATTA TTATTTTCAATC TTCAATTGAG AAGACCTCTCT GTACATGAC TTGGGCAATG  
1020 TTGGCAACAT TGTGTGTAGC GTGATGATCC CTAGACCAAG CTGCAAGCGC GAGGCGCTTCC  
1080 AAGATGATCT GTTCCACATG CTGAGTGAATG CTGAGAGCGC CTACCTGAAAT GGCCTGAAATC  
1140 CTGACAGCTT GGCACCGAGAC TTGGCCCTCT CTGTCTTCTC CGAGTTCCAG GATGTGACAT  
1200 TCTGTGATAT GATGTGTGAG TTGTGTGAGG CACTGTGAGG CAGCTTTTCC TTGGGCACTGA  
1260 CACTTCAATG GAGGCGACAC GCGCGCGGAG AAGATCTGAC GTGCAAGCGC TTGCTGTGAGA  
1320 AAGAGGATCC AAGAGACAC GAAATGTGAT GCGCGAGAGG GACCGCTTCA CCGCGCTGAC  
1380 CATTGACATG GTGTGTGATG GAGAGGCTG CACTGACATTA CTTGCAATC TTGCAACACT  
1440 KACCTTGGAG CAGCTCTGAG AAGGAGCTGAG GCTTACACCC TTGAGCTGAC CAGATTTTCA

1500 GAGGAGACCT GAGGCGCTGAG CAGCTCTTCA GACACTATCT GGTGATCTCA CCGGAGAGAG  
1560 TTGGCGCTCA GAGAGGAGAT CCGCTTCAAC AAGGAGAGAC CAGAGGAGCTC GTTGTATTTT  
1620 GTTACGAGAT AAAATTTTGA GCGAGACCCG AAGTGTCTCC TCTGTCTTTT CTCTGAGTGG  
1661 CATTGTATCT TTGACCGCTT CTTCACCCA GAGCTCTTCA A

## (2) INFORMATION FOR SEQ ID NO: 242:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1146 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

60 TACACAGAAA GCAAGAGATG AGACTCTTT CATTCACTTT TCCTAAGCCC ATTCCTGTCT  
120 GATCTTCCCG CCGCCATCA TACCTCTCC TTCTGTAGAC CTGTCCCGCG TTGGCTGTAA  
180 TGTGTGACTT TACCTGATA TTTCATGAGG AGGATGAAA GCGAGATCA CCGTACCGAG  
240 TTGGACATAT GCGGAGAGGA AAAAGATGAG GATGACAG AGAGCGGTC AGACAGAAAC  
300 AAGAGATTTG GGTGAGAGGA AAAATGTAGG GAGAAATGAG GTTGCAGGCG CTGCAAGCGG  
360 GTTTCAGCAG CAGCTTCCCG CTCCCGGAGC CTTTGCATC CAACTTGTCA GACAGAGGTC  
420 CAGCTCTCTG GTGTGTATCA TGAATTTCT TCAATGATG TTATGCAATG TCTTCCGTAG  
480 GTTAAAGAC CCGTGTGTG CACTGTACA TTGACACCTT AATATATTAAG ATTAATATAT  
540 AATATATAT ATGTAAATTA TGAACATGAG GCGCTTCTCC CCGTCTGGA CCAAGCAGAAA  
600 CTAGACCTTT TGGTGTGAGT AATATGTTTC GTTTGTGTAT TTGTGTGTTT TTGTGAGCTG  
660 TCTTATGTCG TGAATGACA AGTGTAGTC GAAATGCTCT GTATTAAGAA ATTAATGTTT  
720 TAAATTCATA AGGTGTATAT TAAATGTAGC CTGAGACCT CACTTTCAGC TAAATTTTAA  
780 AAGTTGCCA ATTGTGTTTC TTCAATTTA CTGTATCTT TTGTGTACA ATTCATCTCT  
840 TTCTCTCTTT CTCTCTTCC CACTGTGAC CATTGCGCTC TCAATCTCCG TCTTCCCGCGC  
900 TCCCTCTCTC CATTGTGCTC CCGGTGTAT TCTGTGAC TCAATCTCT CTGCGCTCTC  
960 CTGCGCTCTC GATCCCGCTC CCGACATCA CTGCGGAG TTGTGCTTCC GCTCTCTTAT  
1020 CTTGTCTAGT TTGGAAGAG TTTCATCTGA AGTTGTGAG TCTGTGTGAG AGCTTTTCCCG  
1080 ATCTGCTTTC GTTTCGTGTA GATTTGAGCG TTTCCTTTTGA ATTTCATGCT TTCTGACAGAG  
1140 AATTAAAAAA AAAAAGAGAA AAAAAGAGAA AAAAAGAGAA TCAAGGAGAG GCGCGGTAC

CAATTG

1146

5 (2) INFORMATION FOR SEQ ID NO: 243:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1350 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

15 AACCCAGGCG TCGTGGGCA GGGCTGGAG GGCAGAGGCG CGCGAGGCG CAGTTCCAAA 60  
CATTGGCTAG AGCAGAGAGG GCGGAAACCC GTTGGCCGAG CCCAGCGAGC TTGACAAACC 120  
CTTTGACCA CCACAGAGCT ATGAGCCTCC AGCCCCCTGCC CCATTGCCCTC CACCCTCAGC 180  
TCCCTCTGTT CAGCCCTGCA GAAGCTCTAG CCCACAGAA CTTAGAACT ATGGCTGATA 240  
CAGCCTAGAG GCTCTAGCTG CAGCAGCCAC AGCTGAGCTG CTGAGAAAC AGCAGAGCT 300  
CAACCGAAG GCAGAGAGAT TGGACCGAAG GAGTCGAGAG CTGAGAGATG CTGCCCCGCG 360  
RGCAGAGCT ACTGACAGCA ACATTGCGC CCTCTACTCT TCTTTTCTC CAGTTCAAGC 420  
CTGCTTTTTC CAGGACATCT CCATGAGAT CCCCCAGAA TTTCAGAGA CTGTATCCAC 480  
CATGTACTAC CTCTGAGAT GCAGCAGGCT GCTCTCTCTC CTGAGCTTCC TGGCCTGCTT 540  
GGCCAGCTTC TGTGTGAAA CCACATGG GCGAGGCTTT GGGCTTTCTA TCTCTGCGT 600  
CCTGCTTTTC ACTGCTGCT GCTTTGCTG CTGTATCCG CCATGTATA AGGCTTTCCG 660  
GAGTGAGAT TCATTCAATT TCTTCTTTT CTCTTCAAT TTCTTCCTCC AGGATGTCT 720  
CTTTGTCTTC CAGGCGATTG GTATCCAGG TTGGGGATTC AGTGGCTGGA TCTCTGCTCT 780  
GGTGTGCGG AAGGCACAC AGCAGTATCC GTGCTCATGC TACTGCTGCG CCTGCTCTTC 840  
ACTGCGATTG CTGTCTAGG AATTGTCAAT CTGAAACGGA TCCACTCTCT ATACCGCGC 900  
ACAGGTGCCA GCTTTTCAGA GGGCCAGCAA GAATTTGCTG CTGTGTCTTT CTCACAGCT 960  
GGGTGTGAAA CCGCAGCTTG CCATGCGAG CGCTGGGCT CCGTCAAAATG CCTTCCGGCG 1020  
CCGCTGAGCC CTGACTGCGA TCGCTGCGC CTCTACTTTC AGGGAGCTGA CTTAGCTCCC 1080  
GTCTCTAAGG TCTCTGAGAC ATCACTAACT GATGCTCTCT CATTAGTCTT 1140  
CCCAATCTTA TGGCATTGAC TCTGTAACT GACAGGGTTC TGGGAGTTTC ACTGTAGCT 1200  
AGTCCCCCCA TCAGGCGACA CTGCTCCAC CTCTCAGAG CCCCACCCA GCTTCCCTCT 1260  
GCTGTGCGAC GGTGTGCTCT TCGTTTATTT AATATTAAG AAGTGGAC TGGAAAAA 1320  
AAAAAAAA AAAAAAAG GGGGAGCAG 1350

5 (2) INFORMATION FOR SEQ ID NO: 244:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1529 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

15 TCCAGAGGC CGGGGGTTC CAGCTCTGCC TGTAGCAGAG CCTGAGGAG GAGGAGAGAG 60  
AGGATGTGCT GAATATACATC CGGAGATCT TTTTCAGCTA GGGATTAAC TGTGCATCTA 120  
ACTGTCTGCC GAGAGGAGCT GAGAGGAGC TGACCTTTCA CTGTGTGTC TGGGGCCGAC 180  
GCTGTGTGGA ATGGGGCTCT CTGTGCTCTT ACCTTTGTGC CTTCCTGGGC CTGCGAGATT 240  
CAGCTCAGGC CAGAGAGCCC TGGACACTCC GGGCCTTTGG GTGCCGTTCCT GAGTGTGCGG 300  
AAGGCGAGAC TCAAAATGAG ATCCATTTTG ACTCCCTCTG TATGTACTCT GGCCTCTCTT 360  
GGCTCTGAG GCTCTGGAGT CCAATTGTC TCTGTTAGTC AGTGACCGAG TTCACAGGAA 420  
AATATGTCA TGTGTGCTC CACTTACTG GAACCAAGA GACATTAAT TGCAAAGAAA 480  
AGGATCACTG CAGAGTCAC TGGATTTGCT ACAGTTTACT CCGATGTC TCTCTGTGAG 540  
GAGGAAGCTT GTTTCAAAAA TAGTTTCCAT CATTGAGCTA TCATGAGCT CCCACCTCTC 600  
CAGCAGGCTT AGAAGCAA CAGCTGCGC ACAGTTCTCT GCGCTCTCTG GAGGTGTGAG 660  
GCCACAGTGT ATAGACTGGT AAGCCAGACA GGCCTCTCTC CCGAAGTCC TACCTTCTTT 720  
TCACCTGTAC CTGTGTCCC GGGGAGCTAG CTTTAAGCA AGAGGAGAG GAGCCAGAAA 780  
GAGACACTGA GGCAGAGAGA TCACACCGA GTACATGTCT CTGCTCTCTT TTTTCACTGT 840  
GCTTTGGA CAATATATG AATAAATCAC TCCATTAAG GTTTTCCAA ACACAAATCC 900  
TAGAAATPAC AACAATTC CCAATGCTA AGTTGTCTA ATGTCTTTC AGTCTCTGG 960  
TTGGGAAGTG GAGGGTGGCA GGTTTTGT TTGCGCAACC GTCCAGTCTT GTTCAGAGCG 1020  
AGGATTTGGA GTCTCTCAG GTCTCATCAT GGGAGTGAAT TGTACAGCGA CGCCTCTGCC 1080  
CTGTCTGCTT TCAGGTCCAG GGAAGCTTTC AAGCAGTCAA GCTTGTCTTT TGTACCCCAT 1140  
GTGTCTCTTC TTTTGTGAGT CACTCAGAGA TCACCTCTGG ACCTCTGGGG TTGAGATTC 1200  
AGTGTGCTT TATGCGGCG CACTCACTAT GGTGGGCTGA GTGGAAGTC CTTAAGCAATG 1260  
TCCCCAGAG CACTGAGTGT CTGCTCTTTT TAAATGCTC GTTTGTGTGC GTAAAGTTCT 1320  
TGTAGGTTT CATTTTGGCA TTGGCAAT CAGCTCGAAA GTCTGCGCCC ATGACAGCAA 1380

1440  
1500  
1529

10 (2) INFORMATION FOR SEQ ID NO: 245:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1537 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

20 GTCGAGAGTC CCGCGAGACC CCGAGCGACC TTCCCGCGCC CCGCGCGCTCC GAGAGCGAAG  
GAGCGCGCTCG AGAGCTCCAC CTAGCTTCACA GCGTAAAGATC CCGAGCGACA ACTCGGAGTC  
AGCATTCGCT AAGCGCGCAG TCGTTGAGCC TCGTGTATTA ATCTGTAGCC TGTCTGTGAA  
TTCGCGTAAA TTTCACCCCT CAGGTATATC TTCCAGTTAC ACGAGATCCT ATCGAGATCG  
TTTGAGAGAT TATCTATCTC TATCGAGATA TGTTCAGAGAT TTTTGAATTC ATCTAGACAA  
30 GCGAGCTCGC AGTTTGGAAA CTGAAATTTGA AGAGTTTCCA GAGAGCTTGA ATGCTGTGCT  
TACAGAGAT GATCTTTTTC AAGAGCTTGT GAGAGCTATC TATCGAGAGC CCGAGATCAT  
CCGAAATTTTC TCTTAAATCG GAGCTCGCCCT GTGTAAATTC CTGTCCGATC ATCTGACAT  
TACCCGACAG AGTGGGAGCT TCGCGGATTT GCTACTTCAA AGATGTGCGA CTGAATATGA  
AGTAAAGAT GAGCTGCGAA AAGCGAGTGA AGTTACTCGA AAGCATTC ATCGATTTGT  
40 ACTCTTCTCG GCGAGACTTT ATCTTAACCT GCGAGTCCAG GAGAGATTC GAGCGGTTAC  
AAGAGCATAT ATCTTCCAGC TTGCTCTTCC AGAATTCCTG AATCCCTCTG TTTCATATTC  
TATGATGAC AATTATATTT GTTCAGTTAA ATGTATTAAG TTGACAGAT CAGTTTTCGA  
AGATCTTCG AAGGAAAAAG GAAAGATGCA TATGAGAGAA ATTTATCGAA GATTTGAAAA  
50 CCGTCTCTTA GATCGAAGCT GCGATGAGCA TGTAAACAG ATGCTCTGA AGCTTGTAGA  
ACTCGGCTCA AGTAACTCGC GCGAGATCGA TCGACTTCA ACGATTCAG AAGCGACAC  
AAGAAATGAT CTTAAGTACT TTATGATGTA ACGAGATTT TATTCATCTG ATGCTGTCC  
TTTCACTCGA GCTGATTCAG ATTACAGAA GAAATTCGA GATTTACTTG AAGAGAGGAA  
55 CTTTTCCTCA GATTATGAG AAGATCGAC AGATTTATCC GCGGCTCGTC ATCCATATCT  
GATGATATTT GATGATGAGA TCGAGCGAGA GATGAGAGAA GCTTGAAGAA AGTTTGTATT  
60 GAGATCGAGG CTTAAGCGAA AAGCTGAGG TTAAATTTCA GCAATTCAT TTATTAAGC 1260

1120  
1180  
1440  
1500  
1537

15 (2) INFORMATION FOR SEQ ID NO: 246:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 506 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

25 TCGAGATTT GCGCGAGACC CCGCGCGCTG GCGGTCTTA TCGCTTCCGA GAGCTGATC  
AGCGAGCGAG CTGAGAGAGG TTGAGGAGAA GTCTGCTTC TGGCTTCGCA GAGCGGATCG  
30 ATTAAGTCGA GCGGAAATTA AAGCATTCGC CTTCTGCTT CAGTGTGAAA GCGCGAGTGA  
AGATCTTCGC GCTGAGATTT ATCGACTCAC TGTAAACAC AGTATTCAGT CTGATCTGAT  
CTGTCTTCGC ACTGATTCGA GAAAGCGAAA CATTGAGAT TGTGTGAGCG GTGTTCGAC  
35 TTGTGACAG AGTATTCGCT CTTCGCGAGC GCGCGCTTAT TTACGCGAG CTTCCTTCCA  
ATCCGAGCG TCTTACGAG AAAAGCTTC TCGATGAAA AAAAGATTT TTGTATTTT  
40 AATTATCTT TATGTTGAT ACTAGATAT AAGCATATTT CTGATTAAT CCAAAAAAAA  
AAAAAAA AAAAATTT TGGTGG 506

45 (2) INFORMATION FOR SEQ ID NO: 247:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1348 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55 GTCTTCTTT TACTGTTTG AGTGTGAG TGAATGATA GGTATCATG GCGCTTCAG  
ATGAGCGCTT GAGAGCTTC GAGATCTTT AATATTCAG TGGATCTCG GAGCTGCGAG  
60 GCGCTTCGA GCGCGAGCTC TCGAAAAAC TCGAGATGTA TCGCGAGAGT TCGAGAGAGC 180

5 TCTGCTGAG GCGAGGTGC CTCTGACAC TGGTAGTTCT GGGCTGCGA GCGAGGCGG 240  
CTCGGGCTTT CTCTGAATG AACACTGCTC TTGAGCAATT CAACTACTTG TTCTCAAAAC 300  
ATTTCATAT TGAATGGTAG GTTTTCATTA GCATTTGTTTC TTTAAGGCAT GGAAGGGAA 360  
GAATGCTCA GGAATGATG GTTTGTTTCA GTGGATGGG CCGGGTTCT CACTGCTGG 420  
GGCTTCCCTT TGCATGTGCG ACCTTTGTGC AGGGCCACCA GGCAGACTCT TCCACTTTC 480  
TCCACTGAA GCACCAAGGG GCTTGAACG TAAITTCCTT AATCGAGGC ATTTTTTTTG 540  
TCTTAGTATC TTTCACACTT GTCCACGCT CTATATTTTT TAAAGTTCT GTTCTGTGTA 600  
TTAAACGGA ACTAGAGAGA AATAGTTTCT GAAGCAGTT TATGTGCAAG ATCCGCAAG 660  
GAGGTTCGG TAGAGAAAA TAGTAGCTG GTTTAGAAAC TGAAGAGGC AAACAGCCAG 720  
GAGCCTTGG AGAGCAATTT GCGAAGATC TACCTGAGA TAAAGCCTGT CCACTGTCTT 780  
CACCACTGA ATACCAAGCG CTCCAAAGTG TTTTCTCTCT TTCAAAAAA AATTTCCACA 840  
AGCTTTTAAA GGTGCATTTA AGATTCATG TCACTTTTGA ATGAACTGC GGGCCCTGCG 900  
AACTGTACG TGTCTAGAA GCTTCAGTC CTCTGCAATG CATGTGATAC TCACTCTCAT 960  
TTTGTTCCT TCAITTCAT TTTGTCTTT TAGCAGATCT GTCCCTGTG GTGCTGTCTA 1020  
AGAACTGGA CACTTGTGTT TTTGTGTAG ATTGAGCTG GCACTGCAA TCACTTCTT 1080  
TATATGCAA TTAGGACAGA CCACTCTGT GTTCCCTGCT TGTGCTTAA TGAAGTGG 1140  
GGAGGAGGG ATGTCAACCC AATAGTAGC CTTCCCATG GCTTTGCCA GGCAGACAC 1200  
TTCACATCT TTACATGCTT CTGTGTAAT TTAAAGTTTA TGTGTATAA GCGAAGCTGT 1260  
TTCTGTCAA CTGTATATTT TGTAAATAAA TATATGCTA CTTTGAGAT AAAAAAAA 1320  
AAAACTGA GGGGGGCGG GTACCCA 1348

(2) INFORMATION FOR SEQ ID NO: 248:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1766 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

55 GTCCGAAATC GCGAGAGCG CACAGAGGC CACAGAGCA GCGGAGTAA AGGACTTGA 60  
GCGAGCAAT TCCGCAATTA TTCTATTTCC CTTCCCTCTC TCCGCGCGG TATCTCTTTT 120  
CACCTTCTC CCAACTGCG TCGGTASCA TGGCGAGCG TCGCGGCCA CTCAGTCCA 180

5 TTCCATCTCC TGTGTGTTCT TCGGAGCCGA GCTGTCCGG CCGGCGGGG GCGGAGGCCC 240  
AGAGCTGCG CCGGCGCTGG GCGAGAGAG CTCAGCTGCC TCTGTCCGG TCCAGATCT 300  
GATTTCTGG AGAGATGGA AGAGACTGG GTTTGTCTTT GCGACACGC TGAATCTCT 360  
GCTTTCCTG CCACTTTTCA GTGTATCAG TGTGTTTCT TACTCATCC TGGCTTCT 420  
CTCTGTACC ATCAGCTTCA GATCTACAA GTCCGTATC CAACTGTAC AGAATCGA 480  
AGAGGCAAT CCAITCAAAG CTAAGCTGA CTAAGACTT ACTGTGCT CAGAGCTTT 540  
CCTAATTAC ATCAATGCTG CCAATGCTA CATAACAGG GCGCTGAAC TCAATTATCG 600  
TCTCTTCTG GTAGAAGATC TGTGTGACT CTGGAAGTG GCTGTCTTCA TGTGCTGAT 660  
GACTATGTT GGTGCTGTT TTACGGAAT CACCTTTCTA ATTCTTCTG AACTGCTCAT 720  
TTTCACTGC CCAATGTCT ATGAGAGTA CAGAGCCAG ATTGATCACT ATGTTGCGAT 780  
GCGCGAGT CAGAGCAAGT CAATTTGTA AAGATCCAA GCAAACTCC CTGGAATGC 840  
CAAAAAAG CCGAATTAAG TACATGAAA CCGAATGCG AACAGTTCT AAACACCAT 900  
TTAATAGTA TAACTGCTT ACTGTACTA TGAAGGAAA TACTCAGTGT CAGCTTGAAG 960  
CTCATCTCA AGCTTTTCT TTAAITGCT GTTTCTTCC ATCTTTTCC TTTAAGCTC 1020  
AGTATCAGG ACAAATATG ATGAGCTAT AAGAGACTA TCTTAGACT CAGAGAGGA 1080  
AGATATCAA TTCAATGAT AATCAATAC CTTAATGCTG GTAGAGCTT TACTGTAGC 1140  
TTCAAGGCG AAGATTTGA GTTAAGAGG AATATGAAG AACACTCTG GGTCTTCTG 1200  
TCCAGTTTC AGCACTAGTC TTACTCAGT ATCCATATA GTTTGCGCT TAAAGATCA 1260  
TGAATACCT ATGAAAAAT TATTTGGGA CAGGATGAG ATACTTCTT TGGTTTTTT 1320  
TTGAGCCTT CAAATCTAT CTCTCTGCC CACAATGGA GCACTAGCC CTCATCTCC 1380  
TTTTCTTTA TGAATTAAT ATCACTTGA TAAATTAAT ATAGTGTATA GTCAATATC 1440  
CTGATTCOA GAATGCATC TGTATAAAA GAATAGAAAT GGAATGCGG ACTGAGGCG 1500  
AGTCAGCAG CACTGCGG TGGCGTAC TCCCTGCGC ACTATGCCA GCGAAGGAA 1560  
RGCTGCCCA TTGCGGAAG TGTGTTTCT GTCACTGAG ACCGTTCTG AGCATTAAT 1620  
TGAAGATCG TTCCGGAATG TGTGTTCTC CTTCTCCCT GCGCACTCA AGTTAATTA 1680  
ATAGGTTGT ACTTTTCTTA CTAATTAATA AAAAAAAA AACTGAGGG GCGCGGCTA 1740  
CCCAATGCG CCGATATCAT COTAAA 1766

(2) INFORMATION FOR SEQ ID NO: 249:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2664 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

AGTGTCTCTG GAGGAGCGCG AGTAAAGGGA CTTCAGCGAG CGAGTTCCG GATTATCTTA 60  
TTTCCCTCTC CTCTCTCCCG CCCGCTATCT CTTCACACC TTCCTCCAC CTCGCTCCCG 120  
TACATGAGCG GAGCGTCCGC GCGCTCTCAG TCCCATTTCA TCTCTCTGTC GTCTCTCGGA 180  
GCGGACCCCT CCGCGCCCGG CCGCGCGCGG AGCCGAGGAG CTTCCCGCCG CTTGCGGAGC 240  
AAGAGCTGCA GCTCTCTCTG TCGCGTCCAG GATCTGATTT TCTGAGAGAG TCGAGAGAGAG 300  
ACTGGGATTTG TCTTTGGGAG CAGCTCTGATC ATGCTGCTTT CCCGCGGAGC TTTCAGCTGTC 360  
ATCGAGTGTG TTTCTTACCT CAGCTCTGCT CTCTCTCTG TCAAGCTCAG CTTCAGCATC 420  
TCAAGTCTCG TCAATCAGC TGTACAGAG TCAAGAGAG GCGATCTCAT GAAAGCTTAC 480  
CTTCAGCTGAG AGATTACTCT GTCTCTCAGAG GCTTTCTGTA ATTACATGAG TCGTCCGATG 540  
GTTCAGATCA AGAGGAGCCCT GAAACTGATT ATGCTGCTCT TTCCTGGTGA AGATCTGCTT 600  
GATCTCTGAG AGCTGCTGCT CTTCATGTGCG CTGATGACTT AGTGTGTGC TGTTTTTCAC 660  
GGAATCCACC TTCTAATCTCT TCGTGAATCG CTGATTTTCA GTGTCCGAT TGTCTATGAG 720  
AAGTACAGAG CCGAGATTGA TCACTATGTT GCGATCCGCC GAGATCCAG CAGTCTCATT 780  
GTTCAGAGAG TCGAGAGGAA ACTCCCTGCG ATGCCAGAAA AAAAGGAGAG AAGATGATCAT 840  
GGAATCCAGG AATGCAGAGC TTACTATTAAC AGCTTTTAAT AGTTATAGAG TCGTTACTTG 900  
TACTATGAGG GAAATATCAT AGTGTGAGCT TGAAGCTGCA TTCCAGCTTT TTTTCTTAAAT 960  
TTGTGTGTTT CTCCCATCTT TTCCCTTTTA CCGTCAATAT CAGCCAGGAA AATTGATGGA 1020  
CTGATTAAGAG AACTATCTTA GAACTCAGAA GAGGAGAGAA TCAATTTGAT AGGATTAAGTC 1080  
AATATCTTAA TGTGTGTAGA GCGTTTACTT GTAGCTTTGA AGCGGAGAGA TTGAGAGGTAA 1140  
GAGAGAGAAAT GAAAGAGAGC CTCTGGGTGC TTCTGTCCAG TTTCAGACAG TACGTATTAAT 1200  
CAGCTATCCA TTATATGTTT GCGCTTTAGA AGTCAATATT AACTATGAA AAAATTAATT 1260  
GCGAGACAGG GTGTGATAGC TTCTGTGTTT TTTTCTTGA GCGCTCAAT CCGATCTTCC 1320  
TCCCGCAGAA TGTAGACAGC TTAAGCTGAT ACTGCTTTTC TTATATGATT TAACTATCAG 1380  
CTTGATTAAT AACTTATAGG TGAATGATAT AATTCCTGAT TCGAGAGATG CGATCTGATA 1440  
AAAAAGATTA GAAATGAGAA GTGGGAGCTGA GAGGAGCTCA GCAAGCAATC TCGCGTCCCG 1500  
GTCTATCCCT CTGCACTAT CCCGAGGAGAA GGAAGAGCTC GCGCATTTGG GAAATGCTTT 1560  
TCTATCTCAC TCGAGACCGG TTCTGAGCAT TATCTTTGAG AGTGTGTCC GAAATGCTTT 1620

TCCCTCTCT CCCCTCCCA CCGGAAATTT AATTAATAG GTTGTACTTT TCTTACTATA 1680  
AATTAATAT CTGTAATCTC TGTGACTGCG TGTAAATCTG TTAGAGAAA AATTAATCTG 1740  
CAGTGGAGCT CCGAGATTAAT TGAATTTTGG TGAATCTATC TGAATCTGCG GCGGAGCAAT 1800  
CTCAGAGGCT GAAATACAG AAGCTTTTCT TTCTGATCT TTCCGAGAA TGCAAATCTC 1860  
GATTTCCAT TTGGGGGAGG GTTTTTTCT TACATTTGAA TATGATTAAT CAGGAGACTT 1920  
GAAAGAGAAA TCAATCTGCA GTTCTCTCAG GTTCTCATC ATATCATGCA TCCCTTCAGAG 1980  
GGAATATGCA CTGCGAGATT TAAATTAAG GCTATGATAT TGAATGTCC GAAAGTACCG 2040  
CAGCTCAGAA AAGTATGCGA AGGAAATGCT CTAGCTGTCT TCGAAAAAT AGTTAGAGAT 2100  
TTGATAGGAT AAAAGGTACC CTTCAGCTTAC TCGATCTTAAT TTCTTACCC CCGTTGAGAT 2160  
GTTTTACTG GTTTCATGTC CTAGTAGGAA GTGATCTTC CAGTCTCATC CTCTGCGCTC 2220  
CGAGAGATC AGTATGTGTC TTTTGTGGCT TCCCTTCGAA AGGAGCTCT GCGATGAGAG 2280  
TCCAGATCC AGTCTCTTC TTGTGTGCT GCTGTGTTTA GATTAATGAA GAGATCTTGG 2340  
TGCAGACAG GATTTTTT TTCTTTAGA AAAACGTATA GATGAAATAT TACTATGAAA 2400  
ACTGTGTGTA GGTGTGTGTC GGTGCAGAT AAAAATAGAG TACAGCTTGA GAGATCTTGA 2460  
TCTGTGTTCC TGTAAATATT CAAATGATG TGTATTAAT AAAAAGAGAA AAAACAGAAA 2520  
AAAAAGAGAG CCGCTCTAGA GATTCAGAC TTACGTACCC GTTCATGCGA 2580  
GCTTCATGAG TCTTTCTATA GCGGTCCGCC AATTCATAT CAGCGGCGCG TCGGTTTTAA 2640  
AAGGTGTGTC AATGAGGAGAA ATCC 2664

(2) INFORMATION FOR SEQ ID NO: 250:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 865 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

CTGAGAGATG AAGTACAGAA TTCAAGCCAT TTGACCCCGA CCGCTCTATT CTGCAATCC 60  
GAGTCTCCG GATTAGAGTC CCGATTTCTA AGGTGAGAT GGTGTCTCTC GGGATGAGAT 120  
TTGGCGTTTC CTGCGGCTTT TGTGTGATC GGTGTCTCTA GATTAAGATT TTAGGTTTCC 180  
TCCGGGCTTT CCGATCTTTC AACTTAATATC CCGAGTCCAA GATTAAGAGAA GCGCGGAAAC 240  
TGAAGAGCTT GATTAAATAG GTTCATCTAC TGTGCTGTGTC AGTGTCTGCG GCGATGCGAA 300

360 TGTGGGTGAC CTTCGTCTCA GCTTCTCTGC TTTTCCGAGG CCTTCCCGCA CATACCTTGG  
420 GACTAGTCCA GAGCAACTCT TTCCCTCTCT ACTTCCACAT CTCCTATGGC TGTGCTCTCA  
5 TCACTCTCTG CATCTGGCT TCAAGCAATG CTTCGGCTCA GCTCAATTC TGGAGGCCA  
540 GCCAGCTTTA CCTGCTGTTC CTAGGCTTAA CCGTGGCCAC TGTCAAGCC CCGTGGCTGG  
600 AACCCGAC CACAGCTGCC ATGTGGGCC TCCAAAGCCT GAGAGAGAG CAGGCGCTGG  
660 GTGGGAGGT ACCAGGAGC CACAGAGGTC CCGATCCCTA CCGCAGCTG CGAGAGAGG  
720 ACCCAAGTA CATCTCTC CCGAGAMT TCTTCCGCTA CCAATGGCTG TCTTCTCTTT  
15 GCATCTGGG CTGCTCTG AGCAATGGC TGTCTCTGC TGGCTTGC CTGGAAATTA  
780 CGAGCTCTA GCATGGGCC TGCATGCTAA TAAATCTTC TTCAGAAAA AAAAAAAAA  
840 AACTCGAG GGGGCGCGT ACCCA  
865  
20  
25  
30  
35  
40  
45  
50  
55  
60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2082 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

(2) INFORMATION FOR SEQ ID NO: 251:

360 TGGGGGGG ANGGGTGTC TGGCTANGG ATGGCAAT CTGGAAATC TCAATACTT 60  
GCTAGCTGT TTTTCTTTT TTTTCTTACA CCCCCCCC CACCCCCG ACTTCCACA 120  
TGTTCATCA TCTCAGAGA GTTCTCATG TGAAGGTG ATCACTTGG AAGCTGCAT 180  
CATTCACATA TTTTCTCTC TCTTCCCT TCAATCATG AACTGTGT CATTTCTGT 240  
GTGTGTGT GTTTTATTT GTTGTATTT TTTTCTTAA TTTTCTTT AGACTTCT 300  
GTGTGCCA CTTTCTTCC AACTCCACC CTCACTCTT CTCAGCCAT CTCCTCCAG 360  
ATGAAGAAA AAAAAAGCA AGTCTTTT TCTCTCTCT GAGTCTTCA TGTGAGATG 420  
ACCTTCGAAA GGAAGAAAA ATGTCAATG TTTATAGCT GCAGCTGCC GAGTTCATC 480  
GGCTTTTCT TTTACATG TTAATCTAA ATAGAAAA AATGCTCAT GAATTCGA 540  
CAATCAAGC TCAATCAAC TTTGGGTGT GACTGTGTG TTTTGGCTT GTATGCCA 600  
ATCTGAGGT TTAATCTGC ATTAAGAAA CTATCTCTA TCTATCAT TATTTGCTT 660  
GCTACTTGT CTTACACA ATGACTATA ACTGTCTCA AGACTTATG GAAAGGAC 720  
ATTATATTA TAAAAAAA AAGCTGCAT CCGTGCATG TATGTATA TTTTCTTC 780  
CTTTTCTTT CTCTTGGT TGAATGCA GTTTCGAGA CTATATGAT GCATTCATA 840

900 CTTTCTTTT ATTGCTCAT GACTTTTGG AGTTTAAAC AAACAGTCC AACCTTAAAG  
960 CTTCTGCC ATGAATTTT GCATCTGTC CAAACTGCT TTGAGTACT CAGACTTCA  
1020 AACTCCAT GCAGTAGG CATCTCTGT GCAAGTATC CAGAATGGT TACATATTA  
1080 AACTGCAA CATGAGAA CTCTTATGT TTTCTTTTA ATAAAGTCA GCGCCACTT  
1140 TGGGACTTA AATTCTCTA TTCCGAGAA GCAGCTTAA ATTTATTTT TAAAGAGCA  
1200 AACTGCCCA TTAATTTGG TTTGTTTAT TTTATTTTA TATTTTGG CTTTGTCTCA  
1260 TTGCAATG TGAATGCT TGGTTTCTA GTATTAAT TAAATCTAGT TTTTAAATC  
1320 TGTAGGCCA GTTAATATGT ATCTCAGA TAAAGATG TTAATGATA ATTTGAAAG  
1380 GTTAGTCTG TTTAGCTGA GATTTTATA ACCATGATG CACTAAATG TTTACTATG  
1440 TCAATTAAG GGGGTAGAG TTTCCAGGG GACTGTATA AAAAGTAGC TTATACAGA  
1500 TGTCTTCCA ACTTAATAT AATTTGGTA TGTGATCT TGTCTATCC ACTGACTGA  
1560 TTGAAACCA AGTATTAAG AGCGAAGC CCGCTGTTA TATCTTAGG GGTATTTAC  
1620 ATCAAAAT GTATGTTT TTTCTTTTC AAAATTAAG TATTTGGAC TGAATTCAC  
1680 TAAATATA CCGCAGCA TATATACA AAAAAATG CAAACTGTT TAGAACCTA  
1740 AAAAAATTA TCAATATA AAATGCCAT TACTGCAG TTTTAAGATG ATGCAGTTT  
1800 TTTTACGTT GTATGTGT GCAGACTGG ATTTCTGTA ACTTAAAAA AAATCCACAG  
1860 TTTTAAGGC AATATCATG AATGTAT TTAGGGACT GACTCTCT CTTTAAAAAG  
1920 AAATGAAG TAAATCTAC CACAATAT ATAAAAAT CTGTCTAGT ACTTTCTTT  
1980 TACATATTT GCTGTGAAA ATGTGTTAT ATCTTGAAT ACTAATAAC CAGCGCTGT  
2040 GTTCTATGT GCTTCTCTT CATTTCAAT TCTGTGATA TCAAGAAAG AATATCTAC  
2082 AATATAAAC GCATTTT TTTCAAAAA AAAAAAAAA AA

## (2) INFORMATION FOR SEQ ID NO: 252:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1482 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

CAGCAGGCT GCGCGCGGG ACTTCTCTT GCGCTGCTC CCGCGAGCG CTCGCGGCTT 60  
CGCGCGTGG CCGTACGGA GTCTTAGGC AGATGGAGG CTGTGTGAA CTTGTACCA 120

180 GAGGTGATGAA AGCAGCGAGAA TCCCGGAGTC CAGGGCTAAC CTCTGATGGA GTCCCCCTTG  
CTAAATGAGCT GCAATTCCTCT GAGCTAGCTG TACTTCCTGTC TCTCAGCTTGG GCGTCGAGTC  
5 ATGGCTAATC GGAAGCCCTT CCGAGCTCCT GAGCTTCATGA TTGTCTACAA CTTCCTACAGC  
GTGGCAGCTT CCGCTGACAT TGTCTATGAG TTCTGATGCT GCGGCTGGCT GAGCAGCATAT  
ACCTGGGAGCT GTGAGCCCTT GAGCTAATCC AACAGCCCTG AACGACTTGA GATGGCTGAG  
10 GTGGCTGGAC TCTTCCTCTT CTCCAGATTC ATTGAGCTGA TGGACAGCT GATCTTAAAT  
CTCCGAAAGA AAGAGCGGGA GGTGACCTTC CTACATGCTC TCCATGAGTC TTGGCTGCC  
TGGAGCTGCT GGTGGGAGGCT AAGAGTTGCC CCGGAGAGAA TGGGCTCTTT CCAATCGAGTC  
15 ATTAACCTTT CAGTGCATGCT CATTAACTGAC CTGTACTGAG GATTATCTGC CTTTGGGAGCT  
GTGGCAGAC CATTACCTTTC GTGAAAAAG CAGATGAGAG CCAATCAGCT GATCCAGTTT  
GTCCCTGCTCT GACTGCAGAT CTCCAGATGAC TACTTTAATCT CAGAGCTGAA CTACAGAGTAC  
20 CAGATCATTA TTGACCTCAT CTGATGATAT GCGACAGATCT TCTTCATGCT GTTCTCAGAC  
TTCTGGATTC AGCTTTATAC CAGGGGAGAG CAGGCTGCCCC GTGCACTTGA GCAAAATGGA  
25 GCTCCAGATTA TTGCGAAGCT CAGGGCGAAC TGAAGAGCAT GCGCTGATTA GCGCCCGAAC  
TAAAGCTCTC AAGATCTGAC CTTAAGGCGAG TTGCTGTCAG TCCCTCTGCC ACCTAGACAT  
30 GTAGCAGAG CTATGATGCT CAGAGCTGAG CAGAGGAGCTG GCGCTGCCCT CCGCAGAGCT  
GCTCTACAG GACGAGGCTT TTGGTCTCTC ACCGACTTCC CCGGCGAGAC TCCAGGAGATG  
35 TGGCTCATTT GGTCTCTGAC AGCTCAGAGC TGGGGGCTTAA AAGGCTGTTA CAGTTAATTC  
CCCTTCCCTG CATTAAACT TGGAGAGGA GAGCTCAGAG CTGGCCCGAC AAGAGGCTTC  
GTGGCCCTTT TCTTCAGACA GAGAGGTGTA GCAATTAAT CACTGTGAGC CCAATCTGAC  
40 TCTTCAGCC CAGACATGA ACGATGAGCT TTCTGGGCTAA AAGTCAAGGT GGGCGGAGGC  
CTGGAAATAC AAGCTGTGGA GAGTCTTAC TCAACTGTG TTCTAATTA AATGTACAGA  
45 GAAAGCTAAA AAAAAAATA AAAAATCTGA GGGGGGCCCC TA 1482

50 (2) INFORMATION FOR SEQ ID NO: 251:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 834 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

60 GCGACAGAG CCGTGGCG CTGGGCGCT ACGAGTCTT TACGAGAT GAGAGCTTT 60

120 GTGAAGTCTT ACCAGAGAT GATGAGAGAC GCAATGCCCC GAGTCCAGAG CTACCTCTTG  
ATGGGATGCC CATTGCTAAT GAGCTGATTT CTCTGAGCTT AGCTTAACTT CATTCTCTCA  
5 CTGGGAGCTC GATGATGAGC TAAATGGAGG CCGTTCAGAC TCGCTGGCTT CATTGATGTC  
TACAACTTCT CAGTGTGAGC ACTTCCTCTC TACATGCTCT ATGAGTTCCT GATTCCTGGGC  
10 TGGTGGAGA CTAAATCTG GAGCTGTGAC CTTGAGAGCT GAGCTTTGAG GAGATGTGCG  
TGAATGCCCT CTCCAGATAC AGCTGTGAGC AAGGCTTATG TTGTGAGAGC TGAAGAGAGG  
ACTGGAGCTC CCGTCCCGAC AGCTGCTGTA CAGGAGCAGC GATTTGGATT CCTGACCGAC  
15 TTCCCGCGAG CAGCTCGAG GATGGAGCTT CATTGCTCTC TCCGACTTCA GAGCTGGAGG  
CTAAAGAGGC TGTAGATTTA TTTCGCCCTC CTTGCCCTTA AACTTGGAG AGAGAGATTC  
20 AAGGCTGGGC CAGAAAGAG TTCTGGGAGC TTCTTCTCA CAGAGAGAG GTACAGCATTA  
ATGTGACTGT GAGCCAGATC TCACTCTGCC ACCGACAGA CTGAGAGAT AGCTTCTGGG  
25 CCAAGCTCA GGTGGGCGCG GCGGCTGGGA ATAGAGCTTG TGAAGCTGTC TTACTGAGCT  
TGTGCTTTAA TTAAATGTGA CAGAGAGAAC CAGAAAAA AAAAAAATA AAAA 834

30 (2) INFORMATION FOR SEQ ID NO: 254:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1508 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

60 TTGAAGTCTT AAAATTTTGA ATCCAGAAC TCTAAGATCC TGAATGTGA GCTGCTCTC  
ATTCTCCAT GCTCAGCTCT CAGGTGAGC GAGATGTGTA AAGAGCTGGA CCGGCGAGCA  
120 CAGCAGACT TGGAGTGTGA CTTGATCAT GAGCTTGTGAG AAGACTTTGAG CAGATGTGCC  
45 AAGTGTCTG GTCCAGAGC CTTTGCTTTT GTCTGCGAGA CTGTCAATGA GATGATCTGC  
CTTCCAGAG ACGAGTTTC TTGTGATCTC ATGGCGAGTC TGTAACTT ACCAATATAC  
AAGTGTCTTA AGCTGAGAT GCTCTCTGCA AAGAGCTTGA GACAACTT ACTTAAAAAA  
50 GACTAATCTT TGGGCTGAG CAGCTCTGAC CAGAGAGCTG TGAAGCAGAC CATTCAATGCT  
CTTCAGATGAG ACGTACAG GAGTTCAG TTATTGTGA GCAATCAGC TGCAGATAC  
55 AAAAAATCC AAGATGCCAT GAGCAGAGC TCTTCAGACT ACTTAAAGGC TTGAATCTG  
GTGTCTTCC TTGCTCATG AAGAGGAGG TTCAATGGGC ATTGCGAGC CATTGATCT 600



GGGATAGCTT TGGGGGAGG AGAGACTTC CTCTCTGG GACTTCATG CAGGTGAG 660  
TTGCTTACAC CCATATACAG GGAATTCAG ATGTGAGAGA AAGTACAGTA AAGCTATTTA 720  
TCTTATCTTG ACTTTAAGG GAAATATTT CTGAGAGAT TATATTTTC ACCGAGCTT 780  
TAAATCTTC TGTCTCTG ACTGAATGA ACTGATTTG CGAGACAT TTTCTTATGG 840  
AAGGATGAG ATTGCTGAG ACTGCAATG CTTTCTCTG GTTTTATTTA ACAATGACA 900  
AATGAAATTC TTACAGCTTG AAGCAGAG TGTGCCAGA TGTGAAAGAG ACTTTCAGTA 960  
TGAGCCCTTA CTCTCTCTC CAGGAGGA CTTCCTGGG CTCTGGCCA GCTGTCCAGC 1020  
CGAGCCCTGT GTGTGAATG TTTGTAGCT GTGCAATGG GAAAGAGGG GTTTTACAT 1080  
CTCTTAAGG ACTGATGCC AAGCAAGTA GGAATGACTT AACTCTTTA CGGAGCATA 1140  
TTGCTGTACA CATTTACAGA ATGTTCTG AGTGTCTG TGTGATTTT TCATGCTGT 1200  
CATGACCTGA AGGAATTTA TTAGACGTAT AATGTATGTC TGTGTATTTT AACTTGATCA 1260  
TGATCAGTTC TGAGGTGCA CTCTCTACA TACTGTACAT ACTGTGAGC ACTCTTGGA 1320  
GTGCTGAGT CTTTATCAT GCTGTTTAA CTGTGTGGC ACAATCTTC TTGTCCAAT 1380  
AAGATTTAT ATTAGACTT ATAGAGAG ATATATAC TTTTGTATG TTTTGTAGT 1440  
TCTACATA AATGCAATTT GTGACTGTA TTAAAAAAA NTAAAAAAC TGAGGGGGG 1500  
CCGGTAC 1508

(1) INFORMATION FOR SEQ ID NO: 255:  
(A) LENGTH: 2514 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(11) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2514 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(12) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

GAGAGACTCA CACTCTCTTT CCATATAC TGACGATGTA GTGAGATG CAGGGAGAGA 60  
GCACCTACT GTTGTGTGCA GGTTTGTGA TGAATCTCAT AACTTAGAG AGGAATTTAT 120  
AGGCTTCTG CTTATGAG CGAGTCGGA AATTTGGCT GTGAAATTC AACTATGAT 180  
AAGTGAAG TGGGATTTA ATATGAGTA TTTGTGTGG CAGGTTACA TTGCTCTAG 240  
TGGATTTCT TCCAAATGA AAGTTGTGC TTCTAGACTT TTAGGAAT ATGCCAGC 300  
TATCTACACA CTCTGCTTT CTTGTGCTT AATATGTGG TTGCAAAAT CAGTACCTGT 360  
TATGAGATA TCTGTGCT TGGAGCAAT TGAGGAAT TGTCTTTT TCCATGATC 420  
ACCACACTG CTTTATGAC TTGACAGT AATTTCTGTT CTTTTCAGA AAGTAAAGA 480

AAGGCTAAA GAACTGAGG AATCTGCC TTCTCAGTGG ACAGCAGGC ATGATCTTTT 540  
TGAATTTTAA GTGCACTCC TCGAGCACT TGTTTTATGT TTAGATGTA TAAATAGTCA 600  
CAGAAATTT AGTGGAGATA ACTATATAGC TGGCGAGACA TTTGTACTCT CGATGCACT 660  
GTGAGATTTT GATTTCAATG TTACTATGTT TGTTCCTTAA AATGTCTTAT CTTTTCAGAG 720  
AGCCTTTGGG AAAAAGCTCC AGGGCAAC CTCTGATGTC TTCTTTGGG CCGGTAGCTT 780  
GACTGAGTA CTGATTCAC TCAACAGCT GATTTGAAA TATGAAAT TATCATGAT 840  
TTTGTGTGA GAGGCCACA AATTTGCCA CCAACTTCA TATTCUAATG AACTGCTGT 900  
GGAATTCGG CAGAGCTCAC CAGGTAAT TCGAATCTCA GCTAATCTCT GAGATGTAAT 960  
ATAAGAAC CTTAGTGTG CGACAGTGG AGCAGATAT TCGAGAACTT AAGATATAT 1020  
TCTCAGACA GCACCTCAA GCTCTTAAT GCTTATCTCT GGTACCTCTA GTCTGGGAC 1080  
AACTCAATT CATATCCTGG GAGCAACACC ATGCTGACAT GTATAGAACT GACTTACCA 1140  
ATCTTGAC ACCTGTGAGT GAGCTTCAT GTTGGAAAT CAATGAAA CACAGGGGA 1200  
AAGATAGA GCTTCTGTC ACATCTATG AAGCCCTCA CCTGCTGAC ATCAGATTTT 1260  
TTCTTAATGT GTATGATG CTGAGGTGC TGTGTATCT TCTGTGATG AAGTTGAGA 1320  
ATGAGGGTA TGAATGGA CGAAAGCTC TTAAAGATA TTTGAGAAC ACTTTGACAG 1380  
ACCAAGGTC AAGTAACTTG GCTTGTCTTA ACATTAATTT TGTATATAAA CAGAGCTGG 1440  
ATTAAATGT GACACATAT ATTAAGTCT ATACAGTAA GTGAGAGCTT CTAACAGATA 1500  
ATTCCGAAC TGTGGAAAT AACTAGAGA CTTTAAAA TAGGCTTTCT TATATTTGAT 1560  
ATTGAGAGA AAGGCGTA AGTGTATGTA GACCTTAA TCACTAATA TCTTTGCTTA 1620  
TAGGACTCA TTGATACAT TACCATGTA TATCTTACT GTTTAATGG CCTGTGTTG 1680  
AATCTCAG CTTTGAAGAC CTACTGTT TCCAGAGA GAAGTTGAA AGTGCATGT 1740  
TTCTTTTGG GTATCTCTG TGTAGGAC TCTGAAATG TTTCAATTA GTCATTTTAT 1800  
ACATAGCAT TATATCACT GTGATCTCT ACTGTGTGG TGTATGAA TCTTTGAGA 1860  
AATATATTT GAAGAGTGT GCGAGAGG AATACATTTT ATAAATGTT TAGTGAAGC 1920  
CCACATGCA CTTTGAATA ATAGAGTTT TAAGTATGTT AAAAATCTAT ACTGAGCACT 1980  
TACAGAAAT TACCGAGAA AACTTTGTA CTTCAACAAA CAGGATTC AGTGTAGAT 2040  
TTGTCTTCT TGAATTTAA GAAACAAATG ACAAAGTTG AATGGAAG CCTGCTGTG 2100  
TTCCACATCT GGTGCTGTT TACATCTCT TGTGGAGCT ACATCTCTT AAGCTTTTAA 2160  
GAGGTATAT GTTGAACACT TCTGTTTCA GTTGTAGACA GATCAGAGG CCATGATAC 2220  
TGACACTGA TTTGTCTGTT TTTTCTCTCT GTCTTTTCC ATGACTCTTA TATATGCTT 2280

CACTCTGAT TATACGAAA ACTCGAAA CTACAAAT AACTGTGTG GTTATCTAG 2340  
AAAAATGCG AAAATATCG TGTATTTTT GGTAGAGAA ATCAATTTG TATATTTAT 2400  
TTCAATCTA ATAAATGTG AATTTGTTT AAGCTTAGG CACATTAATT TTGTGGGGT 2460  
CAAAATCTC TTGTATTAAT TCTCTTAAC ATTGATGAA CACTCTACA ATTC 2514

## (2) INFORMATION FOR SEQ ID NO: 256:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2157 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

CTGCTTATG AAGCGATGC AGAATTTTG GCTGTGAAT TTCAACTAT CATTACTAG 60  
AAGTGGGAT TAAATTTGA GTATGTCTT GGCAGGCTT AATTTGCTC TATGGATTT 120  
TCTTCGAAA TGAATGTTT TGTCTTACA GTTTTACAGA AATATCCCA AGCATCTAC 180  
ACACTGCTT GTTCTGTGC CTTAATATG TGTGTGCAA AATCACTAC TGTATGCGA 240  
GTATCTGTG CATTCAGAC AATGAGAGA GTTGTGCTT TTTCCTATG ATCCAGCAA 300  
CTGCTTTAG AACTTGACA GGTAAATCT GTTCTTTTC AGACAGTAA AGAAAGGGT 360  
AAGAACTGA AGGAATCTG GATTTCTAG TGAACAGCA GCGATGATC TTTCGAATT 420  
TTATGAGAC TCTTCAGAC ACTGTGTTA TGTTCAGTG GTATTAATG TGAAGAAAT 480  
ATTATGAGA KTAATATAT AACTGCGCA GATTTGTAC TCTGAGTGC AGTGTGAAAT 540  
TTTGATTTCA TTGTACTAT TGTGTGCTT AAAATGTCC TATCTTTAC AAGAGCTTT 600  
GGGAAAAAC TCGAGGGGCA AACCTGAT GTCTCTTTG GGGCGGTAG CTTCAGTGA 660  
GTATGCAAT CACTACAGA AGTGAATGA AATTTTGAH GTTATCTAG AATTTGCTT 720  
TGAAGAACG AGAATTTTG CACCAACT TGAATTTCA ATGAACCTC CTGGAAATT 780  
CGGAGAGCT CAGCAGGTTA ACTTGATTC TCACTAACG TCTGAGATT ACTTAAGA 840  
AAGCTTAAT GTTCCAGAG TGAAGCAAT TATTCAGAA CTTAAGATA TATTTCTGA 900  
AAGAGCTC AAGCTCTTA AATGCTTAT TCTGTACCC TCACTCATG GACAGCTAA 960  
ATTCAATAG TCGAGAGAC ACGATCTGA CATATTAAG AGTCACTAC CGAATCTGA 1020  
CACTGTGCA GCTAGCTTC ATTGTGAGG AATCAATAG AAGACAGGG GAAAGAAAT 1080  
AAGCTTGGG TCGACATCT ATGAAGGCTT CCACTGAGT GACATCAAGT TTTTCTTAA 1140

TGTATGCA TTCTGAGG TCTGTGAT TCTTCTGTG ATGAAGTTC AGATAGCG 1200  
GTATGAATAT GAGCAAAAC GTCTTAACG AATTTGAGG AACCTTTGA CAGACAAAG 1260  
GTCAATGAC TTGCTTTTC TTACATGAA TTTGTATGA AAACAGACG TGAATTAAT 1320  
GCTGACACA TATATTAAC TGTATCAGG TAAATCAGG CTTCATCAG ATTAATCCA 1380  
AACTGCGAA AATCACTAG AACTTTTAA AATAGGCTT TGTATTAAT GATATTTGA 1440  
AGAAAAAGC GTATGTGAT GTAGAGACT TATATCTAA AATATCTGC CTATAGAGAT 1500  
CGATGATTA CATTAAGCAT TGAATTTTA CTTGTTTAA TGGCCCTGT TTGACTCTC 1560  
AAGCTTTGA GACTTACTG TTCTTCGAA AGAAGAGTT GAAATGCCA TGTTCCTTT 1620  
TGTGTATCT CTGTATAGG CACTGTGAA TTGTTTCAAT TAACTCATTT TGAACATAC 1680  
ATTATTAAT ACTGTGATC TCACTGTCT GGGTGTATG AATCTTTGA AGAATTAAT 1740  
TTTGAGAGG TGTGAGAGA AGAATTAAT TTATTAAT GTTGTATGA AGCCCAAAAT 1800  
TAACTTTGA CTATAGAGG TTATTAATAT GTTAATAT TATCACTAC AGTTACAGA 1860  
AATTAAGCA GAAAGCTTG TAACTGACC AAACAGAT TCACTGTAG AATTTGCTT 1920  
TCTGAACTT AAGAAACA ATGCAAACT TGAATGAA AAGCTGTG TTGTTGACA 1980  
TGTGTGCTT GTTATGATC CTGTGTGAG CTTAGATTT CTTAGACTT TTACAGTGA 2040  
TATGTGAAC ACTGTGTTT CATGTGTAG ACAATAGAG AAGCATGAA TACTGACAC 2100  
TGAATGCTT GTTGTGTTT TCTGTCTTT TCGATGACT TTATTAAGT CCGATCTTG 2160  
AATTAATAC AAACCTGCA AAACCTGCA AATAGGCTT GTGTTATC TGAAGAAAT 2220  
TGAATTAAT TGTGTATTT TTGTGTAG AAATCAATT TGTATTAAT TATTTCAATC 2280  
TAAATTAAT GTAAATTTG TTAAAGCTT AGCAGATTA TTGTTGTG GGTCAAAA 2340  
TTCTGTGTA AATCTC 2357

## (2) INFORMATION FOR SEQ ID NO: 257:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 689 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

ACTTCTGCT GCAAAAAGAT GTTCAAGCT TATTTTAAAC TTGCTAGCC CTTCCTTTT 60  
CATTTATGG AGTGAAGTGC AGCTTAAGA AAGACTGTT TTGGAATGG AAGATTAAT 120  
CAAGAACAG GATGTGATG CGAGCTGTC TCTGCTGCT TCGAATATGC TCGACCCGG 180

5 ACCTCTTACT GGCAGAGCT AGAGATATG TGCAGCTACT TCCATCTCTC TCTCTCTCTC 240  
CGATTTTAGC CCAGCAGCAC AGGTTAGGTT CGAGTTTTC TCTCTTTTCA TAGCTGTAG 300  
GCCCTTTCTG CGAATGGTTC TCATCTCTCT TAATCTATT TAAGCTAGT TTTCTCTCAT 360  
GTCTCCAGCC TCCATCACT GCCACCACT CCCACAGAG ATGCCCTCTCT CATCCAGCTG 420  
GGCTTTTGAC TCCCACTGCT TGTACCCCTC TTGTGTGGAC GCCCTCTCTCC CAATACTTTC 480  
AGCAACAGC TTTCATATG GAAGTTGTCA CTGTCAAGGS CTTTACATC AGCAACAGCA 540  
AATCTTACAT GTCTCTAGG GTCTCTGCTC ATTAGATGC AATTAATATG TAAATACATA 600  
AATAACAGCA TAGAGAGAAC GTAACTCTTT ATTCTCAAT ATGATGTCTT ACHTAGAATA 660  
GCCAATTA TTAAGATAG TAAGCAAT 689

## (2) INFORMATION FOR SEQ ID NO: 258:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2177 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

30 TGCAGCCAGC CTTCCGCCA TGTGATGAT CTGTGCTATT CCAGAGAGCC GGCCTATGCC 60  
ATCTTCTTCA TAGCTTTTAC TGTGATAGGG GAGGCCCTCCG GGCCTGTCTCT ATCCCTGTCC 120  
GGCCACCTTT GCGTTGGTAT TGTCTCTGTA CTGGTGGGCC CCTTGACCTT GGCCTGTCTC 180  
TCTTCAAGCA AGCCTGTATC TGAATCACTC GCTGACAGCC ATCAATCTCA GTCACTTCCG 240  
GGCTTACTG ATGAATCTC TCCAGACTC GCTGTCTCCG AGCGGGTGG GAGCCCGGCT 300  
GGCTTTGAAG TCTTATCTC CATGTGGGG GAGGGAGGAG CTTTCCCTCA GGCAGTTGGG 360  
GTGAAGCCC AGAATCTCT GAGGTCTCT CAGAGAGTCC AGCTGGAGAG CTCCCHAGCA 420  
CAGGCCATGA TCGAGAGGT GCGTTCTAT GGCAGTGTTC TGTCTCTCAGC TGAGAGTTT 480  
CAGAGCTCT TCAAGAGCT TCAAGAGAT GTGGTTAAG AGCAGCCGCC GAGGCCCGAG 540  
TACAGCTCT GCTTCTTCCA GAGGAGCCA GTTCTCTCTC GGCAGTAC TACTTTGACT 600  
AGCTGGGGA CCTCATGCC CTGGCAACC TGGTGTCCAT TTGGGTGTTC CTGGTCTGG 660  
ATGAGAGAT GTCTGTCTCC TGAGGGTGAAT GACTTCATCC TGGGGGGTCT CAATCGCTC 720  
TTCAATGTGT ACTAGCTGTT GGAGATCTG GCTCAAGTTC TTTTGGCTTG GGGCTTCCGA 780  
GGGTACTTGT CCTAACCCTCA GCAAGTGTGT TTGAGCGGCC TCTCTCAAGT TTGTCTCTGC 840

5 TGGWKGSM GATCTCAACT CTGGCTGTGT ACCGATTTGCC ACACCCAGCC TGGAGGCCGG 900  
AATATGTGG CCTCTGTGCG CTGTGGACA TGACCCGCAAT ACTGAACATG CTCATGTGT 960  
TCCGCTTCTT CCGTATCATC CCGAGCTGA AGCCGATGGC CGTGTGTGCC AGTACCGTCC 1020  
TGGGCTGTGT GCAAAACATG COTGCGTTTG GCGGATCTCT GGTGTGTGTC TACTAGTAT 1080  
TTGCCATCAT TGGGATCAAC TTGTTTAGAG GCGTCATTGT GCGTCTTCTT GGAACAGACA 1140  
GCGTGGCCC TGCAGATAGG TGGGCCCTCT GTGGAGCTTT CAGAGACTG GATTACTGCG 1200  
CCACACTT CAGTACTTTT GCGGCTGCC TGTCTACTCT GTGGAACCTG ATGCTGTGTA 1260  
ACAATGSCA GGTGTCTCTG GATCTATATC GGCCTACTTA AGGCCCTGTG TCCAGATCT 1320  
ATTTGTATTT GTGGTGGCTG GTGTCTCTG TCATCTGGGT CAACCTGTTT CTGGCCCTGA 1380  
TTCTGGAGA CTCTCTTCA CAGTGGACC CCGCAGCCA CCGTGCAGCC CTTGTCTGGA 1440  
CCCCAGAGC CACTTACCAG ATGACTGTGG AGCTCTCTGT CAGGATATTT CTGAGAGAGC 1500  
CCGGGAGGA TGAAGTCA CAAGAGCTGA GGCAGACCC GACCTGTGCT CTGTGAGGT 1560  
GAGCTCCGG TCTCCATCC CAGCAGGGC GGCAGAGAG AGAGGCTGGC ATACACAGG 1620  
TCCCATCAT GGAAGAGCG GCACTCTGT GGCAGGCCAG GCAAGAGAG ACCTTCTCTC 1680  
TGACGACCA CTAAAGCTGG GACAGAAC CAGTCTCTTG CGTGTGGGCC AACACCAT 1740  
TACAGACAG CTGCTGTGTC TTGAGGGAG GCGCTGTGCC TCCGCTTCTT TTTATAGCTG 1800  
CTTACGTAG AATCTCTTG TCGACTCCAC AGGGAGCTTT CAGACAAAA TCGAAGAGC 1860  
AGCGGCTCC CTTGTCCCTT CGAGCTTCCG TGGTGGCTTT GCTGGCGGCA GCGCTTGGGG 1920  
ACACAGGCC TGACAGGGC CTGACAGGT TACCGCTCAG ACTTCCGGGG CATTCAGCTG 1980  
GGATGATAC TATATCTCC GATTTTACC CAGCAGCACA GGTATAGTTC CAGTTTATAT 2040  
TTCTTCCAT AGCTGTAGG CCGTTTCTG GATGGTTTAT CATCTCTCTT AATCTATAT 2100  
TGGTCAATTT TTCTGTGATG TCCCGAGCT CCGATCATG CACCGCATC CCCACAGAGA 2160  
TGGCTCTCT ATCGAGCTGG GCGTTTACT CCGACACTGT GTACCCCTCT TGTGTGGAG 2220  
CGCTGCTCC AATACCTTCA GCAACAGCT TTCCAAATGG AAGTTGTCA TGTGAGGGC 2280  
TTTACATCA GCAACAGCAA AATCTACATG CTGCTGAGGG TCTGTGCTCA TTAAGTGA 2340  
ATAATATGT AAGTATATTA AAAAAAAAA AAAAAAA 2377

## (2) INFORMATION FOR SEQ ID NO: 259:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1193 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

```
5      TGTGATGCCC GTGCCCCCGC CCGTGGCTTT TGCCCCGCTG GCGGGAAGTT CATTGTGCTG      60
      ATTTCAGAG ACTCCAAAGC GAGGCGGGGG ACTGAGAGTG TGGGTGTGCA GCGCTGTGGC      120
      AGAGGGTTAA CCGGGGTCAA ATGCACGAGT TCTCACTGCG TACAGATTAC CTCGCCCGCG      180
      GCACTGCCCG GAGGATTGTA AGTCCGTAAGC GCTCAAGTTT GTCCGTAATC GAGGAGAGGC      240
      CATGAGGTG CCGCAGCCCG GACCGCGGAG CTTCCTCTGT AGAGCAATTG GCGTAATTTCC      300
      CGGAGTCTTT GCTGCGGAGG CTGTGACTGC GCAATTCGAA GTCCCTTGAG AGCTTCAGAA      360
      GCGCTTCCCG TACTGCCGAG AGCCCTATTG CCGGGAATCT GAGTGGAGCC GCGTCCCGGA      420
      GCGTTTGGC AAGAGCAGAG TGAACACTAG TGTGAATGTA TACCGAATTA AGAGTCCCTT      480
      AAGCCATTTT GTAAATGCAG GAGCTGTAC GCGAAGTCTT TTTAGATTAA AGCTAGCCCT      540
      GCGTGGCTGG TGCGTGTGCG CAAATTTGGA GCGTTGCTGG GAGTCTCTGT AGAGGCGCTG      600
      CTGATGGCAAT TTGAGAGTGA CTCTGTGAG ACTGTTCGAG AAGAGAAACA GAGGAGTCAA      660
      AAGCAGATCC ATGAGCTAAA ACTGAGAGAG TCGAAGGCA GACTAGCAAT TACTGAGCAC      720
      CTCCTGGAGA AAATTTGAAA TACTTTACAG GAGGATGAGC CTGAGAAACA TGTGAGAAA      780
      ATTGAGCAG TGTGAACTCT TGTGAGAAAC CATTGAGTGA TTAGTAAACA AGACAGAGAC      840
      TGAAGATGCT CTGAAGTTGA AACTGACTGG AGAGCTGAGG GAGCTGCCA TGTCCGATGA      900
      ATGCCAGAG AGAGGCGACT CTTCGTGTAG CCGTGTGAGA AATTTAAGTG CTGTGACTGG      960
      TGTGTGCAAT GCGTGTGCTC TGTCTTTTTC TTTTCTTTT TACTTAAGAT GCGGCTGTTC      1020
      TACTGTCACT TTAATTATCC TTAATTTA AAACAACTT ATGTTGTAT TAACTATCA      1080
      ATATATCAAT AATATATAT ATGCAGCCAG CTGATTTTA AGCAGTAAT AATACATTC      1140
      GCAAAAGATT AAGTTGAT TTTACAGTGA AAAAAAAAA AAAAAAAAA AAA      1193
```

(2) INFORMATION FOR SEQ ID NO: 260:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1262 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

```
60      GAAAAACCA AAGATCCAGA CAATCTCTTT GAGCATGAT TGGGAGCTCT CAATATGCT      60
```

```
5      GCAATTAAC GAAAAAGAA AAGACAACT CTCTTAATA ATCTGCCC ATGTTGAG      120
      GCGTGTCT TCGAGCCGA TTGTCAATT CGAAAGCAC TTGTAAAA TGAAGAGCC      180
      AACATAAAC AAGCTAAC GATGTCTCCA ATGTAGACA ATGAAATTA TCAATTAAT      240
      CTTCGGATT TTAAGCGAA GACTACTTGC CATCTCCAG GACATGAAA GACTGAAGAT      300
      ATAAACTTT CAGATCCAG TATCCAGAG GAACTGTG TGCTTAAGA CCGAAGTTCA      360
      ATATGATAT TTTGGTACG TGTCTTTCA GCAATGATA TTTTGTGCA AAGTCTTTC      420
      GTTGACAGC CATTAATGAC AAGCAGAAA AAGATTATC AACCATCTA AAGAGTGA      480
      GAATTTGAT CTTAAGAGC ACTAGTTTG GCAACTTAA GATTTAAGT TAAATTTTAC      540
      ATAGATTTG ACATGATGC AAAAGATAT GAAAAATCT AGATTAAATA GTTAATCTG      600
      GCGCTTTGT TAAACTGAA GATTTGAAA ATGTGTGTC ACTGCTTTC GAGCTATGA      660
      ATATTTTGT GAATGTGAC CATGATTGA TGTGTGATC ATCCATGAG AACCAAAAT      720
      TTAAATCAA AACATGTGT TCAATCAAGT AATTTCTAC ATTTGCAAT ACTATGTGT      780
      ACAGACAGG TGAAGGAAA TGCTGTCTA GCTGCGTGG TATGTTAATA GCGCAATTC      840
      AACAGAGAA AGCCAAATTA GTTTTTCCT TTGAAATTT TTTTAAAT TATTTCAAG      900
      GTCTTTTAT TAAATATAT GTTGCAATG TTAACAATA TGTTGATGT CTTTGAGCC      960
      TAAATGCTTT TTTGTATTC AGAATGTG TACTAATTTT AATTAAATA AATGATCTT      1020
      CCGTTTCCT GTTTAAT TACTTCTC TGCCTTATC TTAATCTGA TGAATGAGAA      1080
      CATATGCAAT CAACTTACA TGTTCATGC TTAGAATAT TTAAGCTTG TGTCTTAAT      1140
      GTTGACAGC TTAAACAGA GTTAGATGTA CTTCAAATAT AACTCATGA TACTTAAGG      1200
      CAATTCCTT GGAAGTGGG CTTTGTGAA GAAAAAAT TACCCAGAG GCAATTCOA      1260
      GT
```

(2) INFORMATION FOR SEQ ID NO: 261:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1179 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

```
55      GCGAACTTT CCGCAGAAC TTGAAACTT GCAAGCGAA AGCTGAATC GTTAAAGTT      60
      GCGTTCAGC GCGGCGCTGG CCGAGAGAG CCGAATTCG GTTCCGAAA CTTTGCCCTT      120
      CAGCGCTGG GCGAGGAGC ATGTCTTCA CCGAGAGAG CCGCCTTGGG CTACAGAGAC      180
```

CTGGACCTCA TCTTCTGCG CCGACTGCG CGGGAGAGGG GAGTTTCHGA CTGTGAGGA 240  
CGTGTGCTG GACTGCTCTGT TGGACTTCTT ACCCGAGGGG GTGACACAGG AGAGCTCAC 300  
ACCACTCAG CTGAGGAG CTATATGCA GAAATGGTT AAGTGTGCA ATGACTCTGA 360  
CCATGAGAT CTATATCCC TGTACAAA CAGTGGCAA AATGTGGAC TGAATTTGT 420  
GGATTCCTC CCGAGCGAGT TTTGATCAG TGTAGTTCT TTTCAATCA AATTAGACT 480  
TCTTCTGTC TTTATGAAT GTTCAGAGA CCCAATGACT GAGCATTTT ACCCCACAT 540  
AATCGGGAG AGGCTTATG GCGATTCCA GGAAGCTTT GATCACCCTT GTACAGAT 600  
CAITCCACG AGGACCCAG AGGAATCG AGGGGAGGC CTGCTTACT ACTCCAACT 660  
CTTGTGAGG GCGTTAGGC CCGCTCTGA TGAATCAG ACCCTTCAA GGTATATGT 720  
TTCCAGGTTT TTATCGACT TCTCAGAT TCGAGCAG CAGAGAAAC TGGAGTCTTA 780  
TTTCCAGAC CACTTGTG GATTCAGCA CCCAAGTAT GAGTATCTA TGACCTTCA 840  
TGGATGTA AATGAGCA CAGTGTGCT GATGGCAT GAAGAGAC AGACTTTAA 900  
CCTTATCAC ATCTGCTTA TCGGTGTG AGCTGACCA AATGTCATC CTATGTGCG 960  
TAATGCTCT TGTATTACC AGCAGCCCC CTATGTAGA GATGCCACT TTACAAATTA 1020  
CTACATTGA CAGTTTACG CAGTATCAC GTGCCAGAA CAGACTACT CCAGTTGCT 1080  
ACCTGCAT TAAATATCT TTAATATGT CTTGTGGGA AGCCATTCA GACAGCAG 1140  
GAGAGAAA AAAAAAAAAA AAAAAAGC 1179

## (2) INFORMATION FOR SEQ ID NO: 262:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1162 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

GCGAACTTT CCCCAGAC TTGAAACTT CGAGCGGAA ACCTTGATC GTTAAAGTT 60  
GGTTGTGAC GGGCCCTCG CCGAGAGAG CCGATTGCG GTTCCCGAA CTTTGGCCT 120  
CAAGGCTCG CGAGCGAGC ATGTCTCCA CCGAGCAG CCGCCCTCG CTACAGAC 180  
CTGACCTCA TCTTCTGCG CGACCTCG CGGGAGGG GAGTTTCAA CTGTGAGGA 240  
CTGTGCTG GACTGCTGT TGGACTTCTT ACCCGAGGG GTGACAAAG AGAGATCAC 300  
ACCACTCAG CTGAGGAG CTATATGCA GAAATGGTT AAGTGTGCA ATGACTCTGA 360

CCATGAGCT CTATATCCC TGTCAACAA CAGTGGCAA AATGTGAAAC TGAATTTCT 420  
GGATTCCTC CCGAGCGAGT TTTGATCAG TGTAGTTCT TTTCAATCA AATTAGACT 480  
TCTTCTGTC TTTATGAAT GTTCAGAGA CCCAATGACT GAGCATTTT ACCCCACAT 540  
AATCGGGAG AGGCTTATG GCGATTCCA GGAAGCTTT GATCACCCTT GTACAGAT 600  
CAITCCACG AGGACCCAG AGGAATCG AGGGGAGGC CTGCTTACT ACTCCAACT 660  
CTTGTGAGG GCGTTAGGC CCGCTCTGA TGAATCAG ACCCTTCAA GGTATATGT 720  
TTCCAGGTTT TTATCGACT TCTCAGAT TCGAGCAG CAGAGAAAC TGGAGTCTTA 780  
TTTCCAGAC CACTTGTG GATTCAGCA CCCAAGTAT GAGTATCTA TGACCTTCA 840  
TGGATGTA AATGAGCA CAGTGTGCT GATGGCAT GAAGAGAC AGACTTTAA 900  
CCTTATCAC ATCTGCTTA TCGGTGTG AGCTGACCA AATGTCATC CTATGTGCG 960  
TAATGCTCT TGTATTACC AGCAGCCCC CTATGTAGA GATGCCACT TTACAAATTA 1020  
CTACATTGA CAGTTTACG CAGTATCAC GTGCCAGAA CAGACTACT CCAGTTGCT 1080  
ACCTGCAT TAAATATCT TTAATATGT CTTGTGGGA AGCCATTCA GACAGCAG 1140  
GAGAGAAA AAAAAAAAAA AG 1162

## (2) INFORMATION FOR SEQ ID NO: 263:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 735 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CGGCTGGGT ATTTGCTCG CACCATGGG CCCAAGGCA AGTGGGAC GAGAGGAG 60  
AAGCATAT TTTAAGAGAA CAGAGACT CTGAAGTTCT ACCTCGGAT CATACTGGG 120  
GCGATGCAA TTTACTGCT TGTAGGTTG GTCTTCTTTT ACTGATGCG CTGATTTGG 180  
GCTGTGGG CTTTGGGCTT TAGTCTGCA GTGTATGGG CCAGCTACCA CTCTATGAG 240  
TGTATGGAC GAGCAGGTT CTCTGAGGA TGGGGCCCTG ATGZATGTTG CACAGAGCTC 300  
AAGATGAGC AGGGATGGC AGAGACTT AAGATGTGA TCTTACTGAC AGCATGTGT 360  
CAGGTCTCA GCTGCTTCT TCTTATGTC TGTGCTTCTT GGTCTCTGG TCCAGCGCG 420  
GCGCTTACC TCTGTGGT GATGTGCTG GCGCCCTGGT TCACTGCGA CAGTGGCAC 480  
CGAGCAGC AGCAATCA GAAGGGCAG CCGGACAGG AGCGCGGCA GATGAAGCG 540  
TTATAGCAT TGCATTTGT GCCACAGGC ACTGGCCCTG GTTGGCTCTG TCAGGCTCA 600

CAAGCCCTCA TCCCTGAGC AATGAGGTC TACATCAGG GCGAAAACA CTCATAGTA 660  
TTGGGTATAC TTAATCTTA TTAGGTCTTT GATTAAATG CTTAATATG GAAAAAATA 720  
AAAAAAAAA ATTTT 735

10 (2) INFORMATION FOR SEQ ID NO: 264:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 783 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AATGATAGA GCTCCGATG TGTGTCTTG TATTTGGGT TTCCGTGGT CTCCGTTT 60  
TCCCGGGCTG GGTATTGCT TCGACCATG GCGCCAAAG GCAAAATGA CACAGAGAG 120  
AAGAACAGA TATTGAGAG GAAACAGAG ACTGTAAAT TCTAACTGG GATGATATG 180  
GGGACCAATG CATTATCTG CCTGTGAGG TTGATCTTCT TTAATCTATC TCCCTCATTT 240  
TGGGCTGGT TGGCTGGGC TTAAATCTG CAGTGTATG GCGAAGTAC CACTGTAGA 300  
GCTGATAGC ACAGACAGG TTCTCTGAG ATGGGGCCT GATGATGGT GGCATGAGC 360  
TCACATAGGA GAGAGGATG GCAAGATGAG TGTCCGCCAC CCGAAGCCA GCGACCTTAA 420  
GCATGTATC CTATGAGAG CATTGTACA GTATCTGAG TCCCTCTCTC TCTATCTG 480  
GTCCCTCTG CTCTGAGTC CAGGCCGGGC CATTATCTC CTGTGGTGA ATGTCTGG 540  
CCCTTGGTTC ACTGAGACA GTGGACCCC AGCACAGAG CACATAGGA AACGGGAGG 600  
CCGACAGAG CCGGGGAGGA TGAACGGTT ATAGCCATTG ACATTTTAC SACTGCGAC 660  
TGGCCCTGG TGGCTCTCT AGGTGTACA GCGCTCTAG CCTGAGACA TGAAGGTCTA 720  
GTCCAGGGC GAAAGAGCT CTGAGATTT GGTATATCT ATACTTAA GGTCTGTGA 780  
ATA 783

50 (2) INFORMATION FOR SEQ ID NO: 265:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1638 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

GGCACAGGC GCGCGAGAG GTGGGGCTG GCGCCCCCG GCGAGAGCTT ACCCTTCCC 60  
GTCCGGGAGC GCGGGGTCTG GTTCGAGGG ATCCCGGAGC AGCGAGAGC GAAAGAGAT 120  
GATTTCCCG GCGCTCCCC CCGATGGA GAAAGAGGA GTATCCGA AATCTGGCT 180  
AAGTCTGGC AAGAGGATG TCTATCTC CATCTCAAT GTTAAGAGT TCAAGACA 240  
GCTCAGTTG GCAAGTAC TGGAAATG TGTGATCTC AGCATTTG ACTCGAAC 300  
TGGAAATG ATCCCTAGA AATTACAGA GAAACAGAG AGCTCGGA AGCTCTCT 360  
CAATCAAT AAGGTAAAC CAGCTTGA TACAGATG CATTATGAC AAACAGATC 420  
AATTTTAA CAGCGGTGA CCAATGAC AATCATCTT AGTATTAAG TGAATTCGA 480  
CCGACAGGA ATGATGAGC AGCGAGTCA GCTTTCTG GAGAGAGC TACAGAGCT 540  
TATGTATCA GATGATAGG AACAAATTA AAAAAAGAG GAATACCA AAGGTCTCA 600  
AGAGTTGGT CAGGTAGCA ATGATGAGC CTTTATCT GCTTTTCCA GTCTTTGCA 660  
CAGAGCTCT GCGCAATCA CAGGGAGCT CTCCGCTCT GTGAAAAG ACCCTGCT 720  
TTGGCTTAC AGCTGTAC CCGTCTGA AGCTTTAT TTGCGAGAG AAGATCTAG 780  
GAAACAGGA GAGGAGTAC AGCAGTAC CAGAAATG GAAAGAGC TGAATCGA 840  
CATCTGTG CAGCTCTG ATACAGAG GATGATAT GTAAATGCA GTGAGATGA 900  
AGCTTAAAG TATGATAGG TAACTTTGA CCACTTTCC CCAAGAGAA ATCTGTA 960  
ATTGACAAA AATTTTCCA CTGGCTTTG CTTGTAGAA AAAAAATGA CCGACACA 1020  
TAGAGCTTT TAAATGACT AACCAATCC TTTTATAG TATTTTGAT GTATATCT 1080  
ATTATCAAA AATCATGTT TATTGAGT CTAGAGCTT AAAATATCT TTTGTATTA 1140  
TGAAGAGGA CCGTAAATG AAGTGAAGT TTTGATCCA GTTGCATCT ACTGAAATG 1200  
TAGCATTAC GTTAAAGTT TTTTCCCC ACAGTTTGA TGAAGAGGA TGAAGATTC 1260  
TAAATTAAT TCCCATTTA AGATTAATG GACTTGAGT TATATACCA TATTTTTATA 1320  
CTTATTTGA AAGGAGAGC TGTACATCT TCCATCTCA CTGAAAGAC AATTAATGA 1380  
TTATATCAC AGATGATG GAAATCTTC TTTGAAAG CACAGAGAT AAGAAACCC 1440  
TGTTAACCT TCTGTGAT TACATCAC TGTGATCCC GGCCTTTAG TTGACATG 1500  
GAGTGGAG GAAAGAGAG CATATTTG CAGTATGAC TATTTGCTT GCGAGTTGT 1560  
GAGAAATCT GCTTTACCA GAATTTTGA GAATTTCTG CTAAATATC ACTTACCTG 1620  
TGTATATTT TTTTCCCT 1638

60

## (2) INFORMATION FOR SEQ ID NO: 266:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1455 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

5 CCGTCCTACT GCGATCGAGG TACCGGGTCC GGAATTCCCA GCGTCGACCC AGCGGTCCCG 60  
TCACTTTGCCA AGCTACTCTGG GAATTAATCT TGAATCTGAC AGTTTGTACT TCAGAACTGG 120  
15 AAGATGATG CCAATGTAAT TACAGAGAA CAAACAGAA CTCGAAAGG ATCTCTCAA 180  
TCAAAATAG GGTAAACAG ACTGATATC AACTTTCCA ATTAGACAAA CAGCATCAT 240  
20 TTTCAACAA CCGGTAAACCA AAGTCAAAA TCATCTTACT ATTAATAGTCA AATCAGACCC 300  
ACAAAGATG AATGACAGC CAGCTCAGCT TTTCTGGAG AAGAGCTTAG AAGGACTTAG 360  
TCCATCAGAT GTACAGAAC AATTAATAA AACATGAAA CTACCAAGG GTCTTCAGG 420  
25 AGTTGGTCCA GGTAGCATG ATGAGACCT TTTATCTGCT GTTGGCAGTG CTTTGCACAC 480  
AAGCTCTCG CCAATACAG GCGAATCTTC CGCTCTGTG GAAAGAGCC CTCCTGTTTG 540  
GCTTACACA TCTCCAGCC TCTCGAAGC TTTTATGTC ACAGATGAG ACATCAGAA 600  
30 ACAGAGAGG CCAATACAGC AAGTACGAA GAATTCGAA GAAGCACTGA TGGCAGCAT 660  
CTTGTCCGA GCTCTGATA CAGAGAGAT GATATGAA ATGACAGTG GAGATGAGC 720  
35 CTAGAGATAT GATCAGTAA CTTTGGACG ACTTTCCCA AGAGAAATTT CTTAGAAAT 780  
GAACAAATAT GTTTCAGTG CTTTTCGCT GTAGAAAAA AATGTACCC GAGCAGATAG 840  
AGCTTTTAA TAGCACTAAC CAATGCTTT TTAGATGAT TTTTGATGA TATATCTATT 900  
40 ATTCAAAAA TCAATGTTAT TTTGATCTCT AGGACTTAA ATTAGTCTTT TGTATATCA 960  
AGCAGACCC TAAGATGAG CTGAGCTTTT GATCCAGCT GCAATCTACT GGAATGTAG 1020  
45 CACTTACGTA AAGCAATTT GTCCCGACA GTTTTAATA GAACAGATCA GGAATCTAA 1080  
ATAATTTCC CAGTTAAGA TTATGTGAC TTCACTGTAT ATAAACATAT TTTTATACT 1140  
TATTGAAGG GGCAGCTGT ACATCTTCC ATCTCAGTG TAAGACAAA TAATGATTA 1200  
50 TATTACAGA CTGATGAAA TTCTTCTGT TGAAGAGAC ACACATAAA GAACCCCTG 1260  
TTAGCCCTTC TCTGATTAC ATTCACTCT GATCCCGGG CTTAGGTTT GACATGGAG 1320  
55 GTGGAGGAA GATAGGCAT ATTTTGCAG TATGACTAT TGCCTCTGG ACCTGTGAG 1380  
GAATCTGCT TTCACGAA TTTCTAAGA TTTCTGCTT AATATACAC TAGCTGTGG 1440  
60 TAATTTTTT TCCCT 1455

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

## (2) INFORMATION FOR SEQ ID NO: 267:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1086 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

5 GCGCTCAGT ACCGTCCCG AATCCCGGG TCAGCCACAG CGTCCTGAC CCAGAGAGAG 60  
CTGCTGTCT ACATCAGCT GGGCTCGAGC GCGTCGCCG GCGGGGGCG CGAGCTGAC 120  
TATGTCTCT TCAGGGCGG CACGTGTGG CATTAATCTT TGTACCCCA GCATCTAGCA 180  
20 GTTGTGCT GTACTAGCA CTCAGAAAT GTGTGTTGAA TGAAGATGC CTGTGACAG 240  
CAGCGGACT TTATCTTTC CTGACCTTG CTCCTATGAC ACACCTCTC CTGACTGCCA 300  
CTGTCACTCC TTGAGAGCAG AACTCTCTTA GCGAACCTGG ATGGGAAACA GCGATGGCA 360  
40 AGGACCTCT GGTGAGCA GCGTACACT TTGATGAAT GAACAACTG AGGCTGTTCG 420  
ACCGAGGT TACCCAGCAG ACGTAGAGC TCAGGAGAA GTGCAAGAC TTTGTGAGCA 480  
50 AATTTGCCA GTTTCAGAA ATAGTTGGTG GTTTAATGCA GCTTGTGAT CAACTTGC 540  
AAGAGCAGA AATGAAAG ATGAAGCCA TCGGTCTCG GAATCTGCT AATCTATAG 600  
60 CAAAGCAG AGAAGCTCA CAGCAGAAC TTCAAGCCT ATAGCAGAA AAGAAATGC 660  
70 ACTTAGAG GTATCGGTT GAATGTGAG CTTTGTGTA AGTAGAGCA GAACAAATG 720  
80 AATTTTGA CCAATTTAT TTTCAGAAT GAAGTGAATA TTTGCTTTT ATAGTAGAA 780  
840 GCGAAACA AAAAAAGCT CTCAAAACA AAAAAACCTC TGTAGCATC CAGCGCTTG 840  
900 ACCAATGAC TATGTCAAA GAGTGGGT GTAGGAATG CAGCCCCCTG AAGACAGAC 900  
TACAGCTG GCGAGCCAG TTTTACATC AGTCACAGC TGTCTCTGT GCGCTGCGAG 960  
1020 TTAGCTCT CACCTCTAT GCTTAGTGG AACTAGCAG TTTGTAACT TTCATCTTT 1020  
1080 TTTTGTAAA TTCAAAAG TTTGGAAGA GAGCAATAA ATTTTGTCT TCNAATGGC 1080  
50 TTGATG 1086

(D) TOPOLOGY: Linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

5 GCGACGGGAG GAGCGGGCT GGTCTGCTG GAGCGGGG GCGCGAGTG GGGCGGGCGA 60  
GGAAACATGA AGCTTGAGT TCCCGCAGT GAGTGATC GAAATACCG TGTCAATAC 120  
AGCGGGGATA TGTGCTGAC ATATGATAG GAGTGGCT TCTTCATAT TCTCTTACTC 180  
AGCATTCCT TCTTGAGTG TCTGTGCT TGCATTGA GAAATATAT ACATATATG 240  
GGATGTAGG TATTGTGCA TGCATGAA GGAACACTC TCGAACTCC TACCGAGGT 300  
15 AAAACAGAG CTCCTTACTC ATTGAGACA ACTGACTAT GAGTACAGT TTACATCTTC 360  
ACGGAAGTT TTGACATTT CTCGATAT TGTATATTT CTGCGAAGTT TGTATACAA 420  
GTATGATCA ACTGACTCA TCTTAAACAC AGCTTCTCT CTGAGTGAC TATTTCCAA 480  
20 AATGCCACA CTAGATGCTG TTGCGATCT TCGAATAT ATGATATGA ATGTTTGAA 540  
ACTGAAAAA AATTTCAGC CTACTGATT TCTTATACG AACGATGCT TATATACAG 600  
CACTGTCTC CTGATATGT GAAATGAA GTATTACAT TCGAACCCA ATGCGCTGTC 660  
25 CTTCAGTGC TGTATTGAG TCGAATATC CATTAATCA TCGCTGTAT TATTAACCT 720  
GGTACATTC TGAAGAGGG GTTAAAGAC AGCGTGGCA GCGCGACTT ATATATTA 780  
30 GGCATACCA GTGAGGTGT AGTATGAAA TTCAAGAAA TAAAGATTT GTAAACACT 840  
AGAACCACT TAACTATTA TGAATGACA TTGTATTAG AAAAGACAT TTCCAGTAT 900  
35 TCGCTGTG TTAATTAAAG CAGACTTACA TGTAAACCG AATCTCTCT ATACAAAGTT 960  
ATTAAAGAT AATTATATA CCGTAAAAA AAAAAAAAA AAA 1003

(2) INFORMATION FOR SEQ ID NO: 269:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1234 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

50 ATTCAGACT ACAATACCA TATTGTGAT GGTGTGTGT TCTACTTCA AAGTATAC 60  
GAAAAATTA TCTTCGACAC AGAGTACTC TGTCAATCA GAAATGAGG TGTAGATG 120  
55 CCAATGTAT CAGTGTGAT TCAATTCAT ACTTCTACA GACGAACAT GAGCTTGA 180  
GTTCGCCACA GTAAATGAA TCGAATAC CGTGATAGA AAGCGGGGG TATGTGCTG 240  
60 AATATGAT TGGAGTGG CTGCTCAT ATGTCTTAC TACATTC CTTCTCAT 300

GTCTGTGTG CTGACACTT AACAAATAT ATACATATC TGGGATGTA CGTATTTT 360  
CATGATGCA AAGACACAC TTTCGAATC CTGACACAG GTAAACAGG GCTCTTACT 420  
5 CATGGAAC AACGACATA TGAATACAG TTACATCTT CAGCGAGTT TTTCAGATT 480  
TTCGATTA TTCTATATT TCTGCACT TTCTAACCA AGTATATCC AACATCTTC 540  
10 ATCTAACCA CAGCTTCTC CTGAGTGTA CTATTCACA AATGCCACA ACTCATGCT 600  
GTTCGATCT TCGAATTA TAAATATTA AATGTTTGA ACTGAAAAA AATTATACA 660  
15 GCTACTGAT TTCTTATAG GAGGAGTGG TTATTAATC GAGCTGTTC TGTATATAG 720  
TGAATGACA AGTATTTACA TTGAGGGCC AATGCTGCT CTTTCAATG CTGTTTGA 780  
GTGCAATTT CGATTAAATG ATGCTCTGT TTATATCAC TGGTACATTT CTGACAGGG 840  
20 GCTTATAG CAGCTGGCG AGCGCCACT TATATGTTA AGGCAATCAC ACTGAGCTG 900  
TATATGATA ATTCAAGAA ATTAAGATT TGTAAAGAC TGAACACAC TTACTTATA 960  
25 ATGATGGCG ATTGTGTTA GAAATGACA TTTCAGTCA TTCACTGTG GTTATTTAA 1020  
GCGACTTAC ATGTAACCG GAATCTCTTC TATPACAGT TATTAAGAT TATTTTAT 1080  
ACCTACATA TTTCATCTT TTATGTAGG YGAGTGTA TCTCTGTT TTATACAC 1140  
30 CAGTCCACA TTATAGGCT ACTTTTGG TTTCGAGGG GTTAAATAG TGTATGTGTC 1200  
AATGAGCCA TTATACAT TATTAGCTT ACAG 1234

(2) INFORMATION FOR SEQ ID NO: 270:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 574 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: Linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

45 NGAATGCGT TTGAGCGCT GTGCTCCG CAAAGTCT TCAAGATAT ATTAAAAA 60  
TATGATGCC CAGAAAGCC TACACACA AATTTACA GAGATTTGG ATAGAAAGG 120  
50 GAGTATGCG GTTCAATGTT TTTAAATCC GAGCTGTGA TAAAGAGAT AAGCTTTA 180  
AAGCTTCA CCGTCTCT GGTATCAT AACCAATTT ACTTGATTA CATGTAAAG 240  
15 AAAAGTCA TGTGCTGTA AATTACACA AGCGTGTTA GATGGAGAC GTGAACTGC 300  
ACTTACACT TGTATGATA CCGTTACT CATGGATGA ATTAATGAT CTGTGATAG 360  
60 CAGTCACT TGTACTCT TTCAATGCT TCCCTTCAG CCGTCCGGG TGTACAGAT 420



ACTCTGAGTA GATTAATTGT CATGCAGGC ATGCAATCAG AATCTACTG AGCCACCAT 480  
CATTTGAAA TAAATACCTC AGTTGTACG GACTTGATGA TCAGATCCA GGCCTACT 540  
TATTATCTAC TCTCAATAA ACCTTTATTA AACT 574

5

10 (2) INFORMATION FOR SEQ ID NO: 271:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1731 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

20 GCTCCAGGT GCGCTCGTG CCGCTGCAGA TCCAGCTCAC TACCTTGGA AATCTTACAC 60  
CTTCAAGCAC TGTGTATTT TCGTGTGATA TCGAGAAAG GTTCAGACCA GCAATCAAGT 120  
ATTTTGGGA TATTATTAC GTGGACAGA GATTGTCCA AGGGGCGCG ATTTTAGAA 180  
TTCTCTTAT TGTAAACAGA CATACCTTA AGGTCTTGG CAGCAGGTT CAGAAATTTG 240  
ATTAAACAG TGTAAAACTG GTACTTCCA AGACCAAGTT TTCAATGTA TTACAGAG 300  
TAGAAGCGC ATTACAGAG ATTCGCGAG TCAGAGAGTT TGTATATTT GAGATGAAA 360  
CTCATGTG CATCCACAA ACTGCGCTGG AGCTAGTTGG CCGAGGATC GAGTTTACA 420  
TTGTCTGA TCCCACTCA TCAGAACGA TGAATGACG GATGTTTGG CTCAGGCTC 480  
TGGCTCAGC CAGGATCAT ATGACACAG AGTGAAGCT GTTCTGCTC AGCTGTAGC 540  
TGATAGAC CATCCAAAT TCAGGAAT TCAGATCTA ATTAGGCA GTCTCTGGA 600  
GTGCGTCTG CTTTCCAAAG TATAGACAT TTCAAGACT GGTATGCTAC TCAGTGTGA 660  
AGGACAGTCA GGTGAGGAC TGTAGGCCA CACAGCTCT TCTTATCTT ACTAGATTA 720  
AAATGTAAG TCMAAAGCG CTCCTTTTT GCGCTCTTA GTCACTTAA CCAGCTAGC 780  
CATTTGAGT CCAGATTTA GTTACAAAG TCAGAGGCTT CCGTGTCTG TTACTTCTT 840  
TTTTTTTAA TGTCTTTTA TTATTAAAA AAAATTACA TGAAGATCC TGTTTGTCT 900  
CTACTGTGA CTCTGATCT ATCTTTCAA AGTGCAGACT CTGTGAGT TTTCTTAAT 960  
TGTCTACTT AAGAAATG AGTACACAC ATGATTTGG CTTTATAT ACTGTAGAT 1020  
GTTATATCT TAATGTGAT GTAGTGTCT TACTTTACG ATTGAITGA ATAGATTTAT 1080  
TGCATGAA TTTACACAA GACCTCTGAA TCATGTACC CACTCCCTC ACAATGTGT 1140  
CCACTTAGT AGTTGCATG ATCTATCGT ACCAATGAT GTTCATAT TACTATCTT 1200  
TCTGAGAT ACTGATTTCT TATTTTGTG ACTATTAATA AATCTCTTT AATATCTCT 1260

CTTTTAACTG AAAGGATG GGTAGAGCG GTTTCCAATG CCATATTAT GTTGGAGCG 1320  
TGTTTTACA TCTTTGAAGT ATGCTTGTCT GAAATATCTTT ACCACATCT TGAATATATA 1380  
TTCTAGTGT CACAGATTT ACGAAAAGA TAAAGCTTGG GTGGATATC ATTTTAAAT 1440  
GTTCAATGTC TGTCTATAT TTTCTCAC CACTCTCAA AATATGTAAT GCAAAAGTC 1500  
TCAGTAATCA TTGTGTAGTA TTAATTTCT GGTCAATGTT TCTCTTCAT AAATTTATTT 1560  
TCATTAAATA CTTTATAGAG GGTTCGAAA TGTTTTCAA ATATGTGAAA TGTGAACTG 1620  
CTGTCTTTA TATTAAAGTA ATTAAGAAA ATGTATTTGT ATTGAATTA TTTTGCTCT 1680  
CACAAGATG CTCATGAGT ATCTTCCAG GAACTCTAAT ATTATTTAA G 1731

5

10

15

(2) INFORMATION FOR SEQ ID NO: 272:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1320 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20

25

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

30 CTCTTTAGGA AGAAGAGTGC AGAGTTGCG GCGGCAGAGC CATCTCTGCC GCTGGCGCCAG 60  
TCACTATGTA GTGGAGCGGC AGACACCCCT CCGCAAAATC TCGAAGTTC TTAGTCTGGA 120  
CTAGGGGAGT AGCCAGAGAC TCTATGTGCG CGGTTTCAAG TCACTGCGCG CTGAGCGAG 180  
CTGCGCTGCG CATCTTTGCG TCTTGTGCG CGGGAGGCT GGTGCAACA GCTGCACAGC 240  
AAGTGCAGA GCAATAAATTT GTTTTGTACT TACTGTGATTA TGAAGTATC AACCATGTTG 300  
TGGTTTTTAT GCTGGGACA ATCCATTTT CTGAGGGAT GGGAGGATCT GTCTACTTTT 360  
CTTATCTGTA TTCAATGGA ATGCCAGTAT GCGAACTCTT AGGATTTGTC AGGAATGGGA 420  
AGCCAAGTGC CATCTTCAA ATTTCAAGTTC TTAAATCTGG AGAAGGAAGC CACATCTCTT 480  
TTGGAGCCAT GAAATATGTC CGAACTCCAT CTGTGCTCA GATTGGAAT TACGTGGAAT 540  
TATTAGAGC TATGGCTCAG CAGACTCTCG TAGGTATGCG TCGTGTATCC TCAGTTGACT 600  
CAATCACTCA GTTCACACA AAGATGTGCG ACAATTTCTA CAATTTGCT TCACTATTTG 660  
CTGTCTCTCA GCGCCAGATG ACACAAGCC CATCTGAAT GTTCATTCGG GCAAAATGTG 720  
TTCTGCAAT GGTATGAGGC AATTTCTGTC TCAATATTA AGGCTTTTTA TACTGATA 780  
TCTATTTGT CTATGAAAT ATCTCTTTT TGCATTTTA ACATATCTT TTATTGTGA 840  
CATCAGACT GCATGCCAT AAGTATGTA CTATAGAGT CTGATGAGA ACAGTCTTGA 900

CCCTAAATAT TTGTATATAT TGTCCCAAT ATGAATTAAT AAAAGACAA AATATATTT 960  
GCTATATTT TACGAGACAC TATTAAGCT TATTAATTT TGTATATTT CATTTAGAG 1020  
TACATCAT GAGATGATT TATACTCAC TGTATACAG AATGCTAAT GATCTATTT 1080  
TCCAACTTC CCGTATTTA TGAATATTC TTTCACATT GATATCTTA GAGATGAG 1140  
GATCCACAG AAGATTTGT TACAGATGAG TACATATGAG TACATAGCA GACATAGAG 1200  
AATCACAG GTTGAATTT GCAATATAG TACATATGAG ACCATGACA AATTTCTTC 1260  
ATCTCATAT CTGATGATAG AGGAAATAGG ATGATATTTG AATTTTCTTA GATCTATCA 1320

## (2) INFORMATION FOR SEQ ID NO: 273:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 515 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

CGCTGAGAG GGGCTGCTT GCCACCTTGG GAGAGGTCTG GAAATGGGCT GCGCTGATG 60  
GCGCTGATGT GAGTAACTTT GCCACATCTT GCTTGGGTGA ACTTTATTTT AGCCCTTCCC 120  
TTGTATATTT TATGAAATAC AAGAGAGGAG TGGGCAAGTC ATGATATGCA GAGATGAGTA 180  
TTCCCAAGAC AAGGACTCTG GAAAGACAT GAGGCAATTC TTTCAGAGAA TGTTCATTAAT 240  
TCAGCCAGAA GCAATCAAT AATTAATGTC TGAATTTGAG CCGACTCTG TGTATAGAAC 300  
CGTTGGCAG ACTCAGAGAG GGCAGAGAC GCTAGATTTT AGTAAACAG GAACTTCABA 360  
GGTATATAG TCCAGAGAG CCCCCGGGCT GTGAAACCC TGTGACTTAA TGTCTAGAGG 420  
AAGCCAGAGT CAATCTTGA CTTTCTTAC ATGAGCTCTG TGTATAGAGG TATATATGCT 480  
GCTATATTT TACAGAGCA GTTCATGTC ATTTGT 515

## (2) INFORMATION FOR SEQ ID NO: 274:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2995 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

TGACACCAT AAGGATTTCA TGAAGAACT AAGAGAGAG GAGATGAGG TTAATCTAC 60

AGTACCAAT GAGGATTTCA TCGTGGCTTT GTGACACCC GATTGATCA GAACTTCA 120  
AAGTGAAT ACCATGATCC GAGCTGGCTT CAGAGAGTC GTGACTTGG CTAAAGACA 180  
CGAGAGCTT CTTGAAACAG CATTCTGACA ACTGTGCTT AATCTGAGC TCTGTAGACA 240  
ACTTGTGCA TGAATCTAT GCGTGAAGAC CAGCTCATT CAGCGGATAT AAGAGCAAT 300  
CCCGAGAC ATTTAGCAGG TTAAAGCTT TATGCTGAG CATGACAT TTAATGAGCA 360  
GATGACTGCC AAAAGAGCTG AGCTGAGAG GATGACAGAG AATAGAAA GAAAAAATAT 420  
AAGGCTACT CAGCGGCTTT TCAATGAGAA ATCCGACAG GAGAGCAGAA AATCTTAGG 480  
TACAGCAAC CTTCTTCCA TCCCATGCT TTGCAAGCTT GAAACAAAA ACCCAAGAT 540  
CAACAGCTT TCTTCCGCTT GGCAGAGCTT GTGCTGTTA GCACTGAGC GGCAGAGCA 600  
ACTGAATAT GCGTGAATC GCGTGAAGCA GTTGAAGCA TTTCACACT TTGACTTTGA 660  
TGTCTGAGG AAAATGTGTA TCGTGTGAT GAAATCAGAA AAGTCTGAG TGAATGATTT 720  
CTTCCGAGGC ATTGAATAGG ACGAGATAG GAGATTAACA GATCAGAGAT TTAATGATAG 780  
CATTTTACA TCCAGATTC CAGACAGAA GTTGAAGAT ACTGCTGAG CTGACATTTT 840  
CGACGAGAT GCGATATGTT AATTTGATTA TTATGATTT GTGCTGCTC TTCAATCCA 900  
CAAGATAGG TATCAGACA CAGCAGATC AATATAAT GAAATAGAG TTACAGAGCA 960  
AGTGTCTAG TCCAAATGTA CAAAAATTT TCAAGTGAAG CAGATGAGG AATATTAAT 1020  
CGATTTCTTC CTGCAATC AATTTGAGAA TTCTAGAGG TTGCGCTGAG TCCATATCT 1080  
GCGCAACCTT GATGTGTC GTTGTGAGG GATGATGTC CTGATGATTA TTTTATATTA 1140  
AAAAATATC GTTCCAGACA CAGATGAGAA CTAACTTTGA ACTTGAAGG AATTTATCC 1200  
TACAGAGAG ACGATCCAG GAAATGACCC CTTTCCGCTC ACGGATGCA AGTTCAGAC 1260  
CATCTTCCAG GGCAGCTTCC CTTACTGTT CAGCTTCAAG TCTATGTCAG AGTAAACACA 1320  
GATGTGATC CATGCAATCT TCTTCAAGCA CCGCAGCAG TGAATCAGG GTTATCCCAT 1380  
CATCAGAGAG CAATTTGAAA CAGACAGAC CAATTTTCA TTCTATGTCAG AATATCCCTG 1440  
CTGTGATAC CAGCATTTAG TTCTTCCCGG GCTTCAAGAG GTCCAGAAC TAAATGAGCA 1500  
GAGCTTAAA AGTGTGAGG TGGCCTTGGG AGTGGGCTG GAGATGAGAG CCGAGATGCA 1560  
GCGAGAGCC GCGAGAGAG TGAAGCTTCT GACTTTGAC TCTTGAAGAC GATTTGCTG 1620  
TTTCAGACT TCGAGAGCA GCGCTGAGG GGCAGAGGC AATCTCAGAG GAGGCTTAAA 1680  
CAAGCTTCC AAAATCCAA CAAATGTTAA GAAAGACAC ACTGCTTCC CAGAGACTCC 1740  
AAGTCCAGG GATTAACAT GTCTAAGAC CCGCAGAGCA CTATCAGATT TGAATCTTC 1800  
TGCATCATT GGTATATAT TTAATGAAA CCGAGAGAT TATATGTTA AAGTGTAAA 1860

AGATTAAATG TGTATTGAG CTGCTTATTT TTTTTCCTTT TTGTAAAGTAA CTATTTTCT 1920  
GTGATATTT ATGTAGATAA AATTGGCTC CTGTTAACC TGTATAGAT GGGGCCAGA 1980  
5 AATGAATAT TTGAGAAAA CAATGAAA GGTCAAGATA CAATGTGTA TTAATAAAA 2040  
AAAGCCAT TAAATAGGTT TCTGGCGGT GAGGGTTGT AAACCTGTT TATCTTTTAT 2100  
GATTATCTT AATGCACTCT TCTTTTAA CTGTACTTC TATCTCAGA AGTAAATTA 2160  
10 TATTAAAAA ATAGAAATCC TCGAGTGT TTAGGAATCT TTTTGTGTA ATCAGGACA 2220  
CTCAATTTAG CAGAACTGA GGGGAGGCT TTTTCCATTG TTTAATGTTT TGTGATTTT 2280  
15 AGCTAAGAG AGGGAACCTC ATCTAAGTAA CATTTCACA TGTATACAG AAAGGAGTT 2340  
CATTGCAATA CTGTCTTTGG ATATGTGTTT AGTACTGGT GTTTAAGGA CAATATAGCT 2400  
CTAGAAATCA GGGTAAATG TAAATGTCA GAAACCTCA GAACATTTGG GGTTTAAAC 2460  
20 TGAATTTTG CTGCTATCC ACCCTAGACA CCAATTAATC TTGTCTTAC CAGGACCCAG 2520  
ACCTTGCCA AGGGATAGGC TGTGTGTGA CATGTGTAAT TTAGATTTG TTTTATCCAC 2580  
25 TTTTATTTCT ATTTATTTAA ATGTGCAATC AACTTCCAC AAATGAGGA ATGAATTTCA 2640  
GAGCTGTTT CTGAATATGT GAGGTGAGA CAAGACGTG CTGCTCTCTT ATGTGATTC 2700  
30 ACCAGCAC TTGTAAACA GTCTGTGTTG CTTTGTGCTT TTTTGTGGC TAAATAGTC 2760  
AACTGACAA GTGACCTGA AAAGGGCTG TCTGGGGCT CTGTTTTTTA GTCTGTGTC 2820  
TTCACTCCG ACCATGTCG TGTGTATTA TCTCATTTGG TTTTATTTGA GGCAGAACT 2880  
35 GAAGCTCTAC CATGTAATG TTTAGAAACA AGACACACTT TGTATTTAA ATTCCTGCA 2940  
GTAAACAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AGGGGGGCC GTTAC 2995

(2) INFORMATION FOR SEQ ID NO: 275:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1990 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

GGAGCCCGG CGSCTCCCGG GCAATGTGAG CAAGCCCTG CTCACGCTG TCTGCCGTCA 60  
ACCCAGAGG ATGAAGTSC TCTGGGCAAT CGCCTTCTG GCTTACGTG CCTCTGTTG 120  
55 GGGCAACTTC GTTATATGA GGTCTATCCA GGAATATGCT GAATCTAAA TTGAAAGCA 180  
GATTGAGAG ATGTTTGAAC CACTTACAGA GAATATCAGA GATTAGAA AAGCTTTAC 240  
60 CCAGAAATAC CCACAGTAA AGTTTTTATC AGAAAGAT CCGAAAGAA TTTTGATTA 300

5 CAGAGGCGC AGKGTTCGTG GCTTCCATC TTAATCTACA AACTCATGAT GGACGGCAC 360  
GAGGTGACC TGGTGGACAA TTTCTTCAGG GCGAGGAGA GAAAGTGA 'GCATGGAATC 420  
GCACATAGA ACTTCGAGTT GATTACCAAC GACGTGTGGG AGCCCTCTTA CATCGAGTTT 480  
GACGATAT ACCATCTGGC ATCTCCAGCC TCCCTTCCAA ACTTACATGTA TAAATCTATC 540  
10 AAGACTTAA AGACCAATAC GATTGGGACA TTAAACATCT TGGGGCTGCG AAAGGAGTC 600  
GTTGCCGTG TGGTCTCTGGC CTCACATCG GAGGTGTATG GAGATCTCTGA AGTCCACCTT 660  
CAAAATGAG ATTACTGGGG CCAAGTGAAT CCAATAGAC CTGGGGCTTG CTACATGAA 720  
15 GGCAAAGTG TTCCAGAGAC CATGTCTTAT GCTTACATGA AGCAGGAGG COTGTGAGTG 780  
CGAGTGCCA GAATCTTCAA CACCTTGGG CCAAGCAATC ACATGAACGA TGGGGAGTA 840  
GTCAAGACT TCATCTCTCA GGGCTTCCAG GGGGAGCCAC TCACGATATA CGATTCGGG 900  
20 TCTCAGACA GGGGTTCOA GTAGCTCAGC GATCTAGTGA ATGGCTCTCT GGTCTCTCATG 960  
AACGCAAG TCAGCAGGCC GGTCAACTG GGAAGCCAG AAGAACAC ACCTCTAGAA 1020  
25 TTTGCTCAT TAAATTAATA CCTTGTGTTT AGCGAAGTG AAATTCAGTT TCTCTCCAA 1080  
GCCAGAGTG ACCACAGAA AAGAAACCA GACATCAAAA AAGCAAGCT GATGCTGGG 1140  
30 TGGAGCCCG TGGTCCCGCT GAGGAGGT TTAACCAAG CAATTCACTA CTTCCGTAA 1200  
GAACTGAGT ACCAGGCCAA TAACTAGTAC ATCCCAAC CAAGCCCTC CAGAATAAG 1260  
AAAGAGGA CTGCGACAG CTGACTCTCT CACTTTTAGG ACACAGACT ACCATGTAC 1320  
35 ACTTCAGGG ATGTATTTT GCTTTTTTTT TGTGTGTTT TAAAGAAAGA CTTTAAACG 1380  
TGTATGAG AACAACTGG AATTCATTC TGAAGCTTCC TTAATGAA TCGATGTGCC 1440  
40 TAAAGCTCC CCGAATAAA CTGCAATTT TCCCTTGCAC TTTTGAATC TCTCTTTTA 1500  
TGTAAATAG CTTAGATCA TCTCTCGTA TTTTCAAGTT TTTTATCTT GCTGTGAGG 1560  
45 CATATGTTG GACTGTGTT GACAGTTTA TTTACTGTT TCTTGTGAA GCTGAAAGG 1620  
AAGTTAAGC GGGACAAAA ATGCCAATTT TATTTATAAA AGTGGTACT TAAATATGA 1680  
GTCTTATAC TATGATAAA GAAATTCCT AGCAGTATG TCAGGTGCTG GTCCCGGCG 1740  
50 ATTGATTTTA GGGCAGTAA AAGATTCGT TGTGAGACT TTAATGTTCT CTTTAAATTC 1800  
AGAGTTTTTC CAGGTCTAC TTTTGAAGTG CAAACTTGAC TTTGAATAT TCTGTGTTG 1860  
55 CATGATCAG GATATTTGAA ATCACTACTG TGTTTGCTG GATATCTGG GCGGGGGCAG 1920  
GTTGGGGGGC ACAAGTTAA CATATCTCTG GTTAACCATG GTTAATATAT CTATTTTAT 1980  
AATATTTGA 1990

## (2) INFORMATION FOR SEQ ID NO: 276:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2436 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

10 AACCTCCCTT AGCTCCAG GGTAAACCG GTGAGCTT AAAACAGAA GAGACAGAA 60  
15 TTTAAAGTCC GTTCATTTGA AATTACAAA CATTTCAT GTTTTATCA AGGATCTCTT 120  
CGACATCCCT CCTCTTTATA AGAGCAGAGT AACATATCC TGAACCTTG TACAAAAGT 180  
TGAATTCGG CAAAGAGAG CAGATGAGA TACAGTTCA GATTACACG CCAAGAAATC 240  
TTGACAGAG CAAAGAGAG ATCCAGGCA GATTATTAAC CCTCCGAGTG GAAATATGAG 300  
CAAGAAATTC GGCACCTTAA ATTATGACA GAGAGAGCC TTGAGGGGAA GTTAGAGTGG 360  
20 CCGAGGTGG GGCACAGAG GAAATGTTAG TCGGGAGAG CTCTACTTAA TAAAGATCA 420  
GCATTTCTA GCATTCAT GACCTATA TACTTAAT CTACTCTCA TTGATTCGC 480  
GGGAGTCC CTTAAACG ACTGCTCT TCACTTAAA ACTTAATGT CTTTACTCT 540  
30 TTGTATGAT GACTACTT TTGTAAGAG CAGGTGTGG TCGAGGTAA AACACAGTG 600  
ATTATTTGG AGCTTGTG TCAATCTG ACTGTGTTT GCAATTCAT TATTCAGCT 660  
35 TCAATTTGG AATCTTTAA AACTTTGAT AATTGTTGT TGAAGCTGT TCAATTTAA 720  
ATTATTTGA ATTCAGTCA GTTCTGCTA TAAAGATCA CAGTTTGAAG AGCTTCTGA 780  
TTTTCTCTT CCTCTTAT TTTTACCA ATTATGAG AAGATTAAG GTCAATCTA 840  
40 AACTTTGTT TTAATGTTT AATAAACGT CTGGGAGTG GTGAGCATT CCTACTAGT 900  
GTCTAAAG CAAATACTT ACATAGCTT CTTAATATAT AAGATTTACA TTACATTTT 960  
45 AAGAGAAAT AGTGCCTT GCAAGGCTA CTTAATCTT TTCCATATAT TGTATTTACA 1020  
ACTTTGAAT ATGAAATCT ACTATTTGA TGAATATGG TATGTATAT ATACATCAT 1080  
ACATAGCAT AATATGAT GTGTGTGTG ATATATATAT ATATGATGC TGTAAATCT 1140  
50 GACTACGAA CATAATGAC TTTTAAAT CAGAGACGG GTAGTCTAC AGGTGATTA 1200  
TCTTTTGG GCTGATCG TTATTAAT GTTATTTAG TTTTACTCC CAGTACAG 1260  
GGATTTTAA TTAGTTCAC TTACATAT ATTTATAT AAAATTTGA ATAAAGGTG 1320  
55 CATTTTCAA GATAATGG AATGCTGT GTTAAATTA CACCAAAAT ACTGATCTG 1380  
ATGTATATAC AGGTCTTAC AAGAAAGAT GTTAAATTT ACAATTTGGA GATTATTA 1440

## (2) INFORMATION FOR SEQ ID NO: 277:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 782 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

45 GGCATCAT TCTCCAGC TTCTGTCC CCAATATAG TTTATTTGT TGTCTGAGC 60  
50 TCACTGTGG CTTTATTA AATCTTAG GGGCTTAG AAGCAATTC TACTGAGAG 120  
GTTTGAAGC ACTTAGACA GGGCCCCA TCCCACTT GGGAGTTTG GTGAGAGAG 180  
GACTTTGG GGAATGAG CAGACAGAT AAGAGAGCT GACTGTGAA GAGAGAGTG 240  
55 TACAGATTT AGTATCCCA AAATGGTTA TATCCCTCC CCTTTACAT CAGAACTTG 300  
TGAATGGGA AATCAGACA AGAGGGGAT CAAAGTAGC TGAATTCACA TCTTCCAG 360  
60 GAGGGGAAA GTTGAAGAT AACCCGGGT GAGAGTGG TGAAGAGCC TGTTAGAGT 420

505

G

961

5 TGTGGCTGAT CCGTCTCTGG TATTAGTTTT TCCCTGGGA GAGGAGGC CTAGGAGAG 480  
GGGACTGACG GGTCCCGCG GGAATTTTC TCCCTCCCTT GATGAGGGA GAGGAGCT 540  
GGCTGCCAAG CCGCTCCCTC AGGAATGCG CTGCGCCAGG AATGCCACC ACAATAGCC 600  
TCTCTTTTT TTCTAGTCAA ACTCTGTTT ATTCTTGGC TTGCTCCCT CTTCTCTCC 660  
CTCTCAACT TTACTTCTGA TTCTCTTTC ATGGAATTTG GATGAGGT TAACTACAA 720  
CAGTCCGCC ACACCAAGT CTTCAGGAA AAAATACAA AGAATTTAA CAAAAAAA 780  
AA 782

(2) INFORMATION FOR SEQ ID NO: 278:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 961 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

30 GAGTTCCGCG TCGAGACCG TCGTCTGGC CGCGGCTTC ACATGGCT CGCGAGCT 60  
GGACTACACC ATGAGATCC CGATCAGGC CTCTGGAGC CAGAGACAA GCGCCAGCC 120  
AGGTGGGAG GAGCGAGAA CTGGCAGCC TGTGTGATT CTTTGGCT GGGTGGCTG 180  
CAGGACAAG AACCTTCCA AGTACAGTC CATCTACCAC AAAAGGGCT GCATGTAAT 240  
CCGATACAA GCGCCGTGCG ACATGGCTT CTCTCCGAG TCACTGGTA TCCCTTCACT 300  
TGTGTTTTG GCGCAGAGC TCGTCAGCT GCTCTTTGAT TATGAGTTG AGAGAGCC 360  
CTCTCTTTC CATCTTCA CGACGGTGG CGTCACTCG TACCGCTAG TCGTCAGCT 420  
CCTCGAGCC CTGCGCTTCT GCGCGCTCG TGTGTGGGC ACATCTTTG ACAGCGCTC 480  
TGTGTAGAGC AACCTGTAG GGGCTCTGG GCGCTGTGCA GCGATCTGG AGCGCGGGC 540  
CGCATGCTG CGCGTGTGC TGTGTGTGC CTTTGGCTG GTGTGCTCC TGTTCAGCT 600  
CCTCTGCT CCTCATCAG CCGCTTCCA CACCGACTTC TATCAGAGC TACAGAGCC 660  
GGGCTCTGCG TGGCGCGAGC TCTACTCTTA YTCAGGGCT GAGGAGTAG TCGTGGCAG 720  
AGAGTAGAA CGCATGTGG AGCGAGCTT GCGCGCGCG GTCTGGCGC GTTCTGTGA 780  
55 TTTCGTGTA TCTGCACAG TCAGCCACT CCGTACTAC CCGTACTACT ACAGAGCT 840  
CTGTGTGAG TTCAAGGCA ACTGGCTCG CTCTGAGGC CATCTGCA TCTCACTCT 900  
GCTCGAAA TAAATGCTG ACATGCCC AAAAAAAAAA AAAAAAAAAA ACTCGAGGG 960

506

G

961

5 (2) INFORMATION FOR SEQ ID NO: 279:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1228 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

15 CGCGCTTTC AGTTGGTCT CCGTGTATC GCGCAAGCC AGTAGGGGA TTGGCTTCC 60  
TCCAGTGGCA GCGCTATCAG ATTGGATAT GTCTTCATA TTGATTGGA TTACAGTGG 120  
20 TTTCACAGT GTCTACAGT TTTAGGANT ATATAGAAA ACTGTAAAC TGTATTTCT 180  
TGGATTGGT AATCGAGAA AAGCAATTT GTTACAGTG CTAAAGATG ACAGCTTGG 240  
ACACATGTC CCACATATC ATCCACTTC CGAGAACTG ACCATGCTG GCATGACTT 300  
25 TACAATTTT GATCTGGTG GACATGTCA AGCTCGAAG GTGTGAAA ACTACCTTCC 360  
TCTATCATT GCGATGTAAT TTCTGTGGA TTGTGAGAC CACGAAGGC TGTAGAGTC 420  
30 AAAAGAGAA CTGTCTCAC TAAAGACA TGAACCATT GCTATGTGC CTATCTGAT 480  
TCTTGGGAT AAGATCGAC GACCTGAGC CATCAGTGA GAGAGTTC GAGAGATGT 540  
TGTATTAT GTTCAGCAA CAGGAAGGG GAGTATATCT CTGAAGAGC TGAATGCCG 600  
35 ACCCTTAGA GTTTTCATGT GTAGTGTCT CAAGAGCAA GTTACGGAG AAGCTTCCG 660  
CTGATGGCA CAGTACATG ATTAACAAA ACTCAGATT GTTCCAGTC TCAAGCTTCA 720  
40 GCGTACTCA GAGATTTGAT TCGTCAACT GCATACCTG AATTCATAG ACTTTTCTG 780  
GTTATAAAC AGATGTTTT TAGATTATA ATTTAAATC ACTTATATT CAATGAGAT 840  
45 TGAAGACTA TTCAAGTAG TTTCAGTATC ACATGTTAG CTTTCTAAT CCAATAAGT 900  
ACTTGTGTT TACAGTTAT ATCTGACAT CAGCCGAGG CCAATTGTA AGAGCACTT 960  
TCCAGGATA CATTGAGC ACTTTTAAAC ACATGAAAC TATAAACAT ATTAAAGAG 1020  
50 TCATCATCT AATTTTTTA TGTACTTTC TCGAAGTCT TTTTAATTT TAGATTAT 1080  
GTCCACCTT CTTAGTGTGA CAGTAAATA TTAGCTTATT CAATGATG ATGATGCTT 1140  
ACAGTTTCA ATACTTTTT TTCTATGCA AAGTCAATC AATAAACA ACTCTAATG 1200  
55 TTGCGAAA AAAAAAAAAA AANTGCA 1228

(2) INFORMATION FOR SEQ ID NO: 280:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1327 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

10 TCTGGGCTCT CCGGACAGCT GAGCAGCTTG ATGAGAGCCA GCGTCTGAT GCGGCACTTG 60  
GGGCGGTGCA GAGATCTGTG GCGCGCTTCC GCGAGGGGCGS CCGAGACCGG TTTACAGTGA 120  
TTTCTGATTT TTAGATGACC TTGAGCAGCT TTAGATTTAA TCGAAAGCCA TGCGCTTCTT 180  
CTTACATAT TCTGATTAAT AACAGATCA TCAATGAGGA GTGTGGAAA GAGCTGACAG 240  
CGCTCTTCA CCAATTTAC CCAATTTAGA TCGACCCACA CCGGACCTTC AACGAGAAC 300  
TACCTCATAT GGTGGAATGG TGGACGAAAG CCGACAAATC CTATGTGAG CAGAGATATC 360  
AGAGATTCA GATGCGCCAG GTGTGTAGAG ACTCGAATCC AATGTCTAG GAGGATTTAA 420  
AGACCTTTT CAGACAGCTC TACCATTAACA ACATTCCTCT TTTCACTTTT TCTGCGGGCA 480  
TTGGTGAAT CTTGAGGAAA ATTATCCGAC AGATGAAGT GTTCCACCTC AACATCCACA 540  
TCTGTCTTAA CTACATGAT TTTATAGGAG ATGGTTTCT CAGAGATTT AACGCGCAGC 600  
TGTATACAA ATACAGAGAG AACAGCTTG TGTGTAGAAA CTGTGGTTAC TTCAGACAC 660  
TTGAGGCGAA AACCAATTC ATCTCTCTGG GAGACTTAT CCGGACCTTC ACAGAGCGG 720  
ATGGGGTTCC TGTGTGAGAG AACATTTCA AATTTGGCTT CTGATTTAC AAGGTGAGAG 780  
AAGCGGAGGA AGCGTAAGT GAGCTCTTAT GACATCTGC TGGAGAGGA CAGAGCTTG 840  
GATGTGTCA ACGGCTTACT GCGACAGATC CTGTCCGAG GGGTCCAGC TGGAGATCA 900  
AGCGCCCTGA AAGCGAGGAC TCCAAAGACC CTCGAGGACC GTGTGAGGA GGGGCGCTTC 960  
CCGAGATCT GCTCCCCCTT GAAACAGAG CAGAGCGAGG GTGGCAGACA GTGGCTGGT 1020  
CGTTCCGAGC CGCTCCCTCC TCTTTCTCT GAGCAGCTTC ATCCAGAGAG GCTTGAAGGA 1080  
AACCGCCAT GTGGAGGAGC ACGAGCAGTG TTTCTGTGA ACTTGGAGC AACAGATTC 1140  
AGTGTCTAG GATTTCTTCT CTGAGAGAT TTTCTTGAAT ATTATTAC ATGGAGATTT 1200  
CAAAAGAAA TAAATGTGCA AACAGATCA AATTTTAAA ATGAAAAAA AGCTGCTTG 1260  
ATTGAGGAGA TGTGGCTGGG GTTAAAGCT GAGCTCTTG CGCTGGGAGC AACATGAGAT 1320  
CTTCTAG 1327

(2) INFORMATION FOR SEQ ID NO: 281:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 799 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

10 TCACTCTGCC TACAGCTTGG ACTGAGAGG ACTGCGCCTT CAGAGCTTAC TGTGAGAGG 60  
TGTATTTAC AACCTGAGG AACCTTACAG CAGAGCTTGG GGTGTCTCCC GTTACATGT 120  
GAGTTTGGCT GAGCTTAAT TCCAGACTC CTGTTTTTTT GAGAGAGACT GAGAGAGACA 180  
TAACTCTGC CTCTGAGACC CAGCTTGGG CTGAGTCTC TGTAAATACA TTAATCTTGG 240  
ATGTGGGTT ATTCCAGACA AAGCAATTC AAGTCTCTGT AACGATTTTG TACATATTTA 300  
TAAATATTA TCCAGAAAT GTTCCAAAT TCCAGTCTT TTGCTCTGGG TACAGCCACA 360  
CGCTTACAC CCGACTTGG ACTTGAAGAC CTACATGAG GAGCTTTCC AGGTCTCT 420  
AAGAGAGAT TCTGTGATTT AGCTGTACAC GAGCATTTT TCCAGAGAT CAGAGAGACA 480  
TTGCTGCTT GATGGGTTAG AAGTTGGTT TGTCTTGT TTGAGCCCAA TATGTAGAAA 540  
ACATTTAAA CAGTTTACAC CTTTATAGG GTATGATTT TCGAAAGCA CCAATTCATT 600  
TTGTGATTT TATGTCTG TGCTTTTAA ATCATATTA CAGATTAAT ACCTTTTCT 660  
CAATTTGCT TCCAGAAAC ATACTTAG TGTTTTTGT TTTGTTTTG TTTTTTTT 720  
TTTTTTTTT CTTTATAGAG AAAAAATTA ATGTCTCAT TTTAAATAC AAAAAATGA 780  
CAAAAAAT CAGCGGGG 799

(2) INFORMATION FOR SEQ ID NO: 282:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

50 AAGAGCTTA AACATCATGA GCTTGAACAT GAGCAGAGC CTACTTGGC CAGCCAGATG 60  
GCTAGACCT TCCGTACTT CCGAGATGGA TGGGTCTCT ACTAGAACCA GCTGTCTTT 120  
CTGGCTGCA TGGCTTTC TTCTTTAT ATGAGCTGCC TGGCTTTTAA CTGCATACCC 180  
AAGGTTAGC CTTACACTCA GAGACTAGAT GGTTCGATCC TCAATATTT GATGGAGACA 240  
TCACTATTA CTGAATTAAT GCGAACTTA GCTTTTACT GCTTACCTGC AAAATGTGAT 300

TTGGTGGCA CAGGTCTGAT CTGAGGNTG GCACAGCTTT CTTGTTGAT CTTGTGTGTG 360  
ATCTCTGAT TCATGCTGG AAGCCCGCTG GACTGTGCG TTTCCTCTT TGAGATATC 420  
5 CGATCAGGT TCATTCAAG AGATCAATT ACACCTACCA AGTATCTGA ANTACAACT 480  
GAATATACA TGTCTATAG GTCTAATCT CTTATATATG TCCGGAGAC AGTCTGAA 540  
TCGTGCGCA TAACTCTGT CAGTCTGCTG TTTCAGGGG TCATCTGTC TAGATCTGT 600  
CTTTGCTCT TCGATTAC TGTACACAG TTGCTCGAG AATATGTAT TGAATCTGAA 660  
AGAGGCATTA TAAATGTGT AGGAACCTCC ATGAATATC TTTCTGATCT TCTGATTTT 720  
15 ATCATGTCA TCTTGCTCC AATCTCTGA GCTTTGCTT TCTGTGATT GATTTCAGTC 780  
TCTTTGTGG CAATGGGCA CATATGTAT TTCCGATTTG CCAAAATAC TCTGGAAAC 840  
AAGCTCTTG CTTCGGTCC TGAATGAAA GAATTTAGA AGGAATACA AGGAATACA 900  
TCGTCTCTT GACACAGTTT AACTGTCTT ATCTCTTAC TAGATTATAT AGACACATG 960  
TCTTAATTT GTACTGACA ATCCATPAA ATGCTGGGT GTTTGTCTT GTTTTACCA 1020  
25 CAGCTGTCC TTGAGACTA AAGCTGTTT AGGAACCTA AGTCAGAGA AATTAAGTA 1080  
TTAATTTCC TTATGTTAG GCATGAAA AAAATTTGAA AAAAAACT CAGTTTAAT 1140  
ACGGAGACTA TAAATCAATC ACTGAATCC CTAATTTCT ATGATGATG ACAAATCTAC 1200  
30 GTAAAGACT GTTATGAC GTGAATCAG TTATCAATTT AGCAATCTT ATCTGACTA 1260  
GAATTCAGAT ATCTAGTTT TGTGMAAC TCACTCTGT TCAAGACTAG CTAATTTAT 1320  
35 TTTTTCATC TTATTTATT TTAAGACA ATCTTCAG TATGAGACT AATTTTAT 1380  
AATAATATT ATCTTATG ATCTATTA TCTTAAGTA TTATCATGA TGTGAAAA 1440  
CAAAACACT AACTAGATT CTCTATAG GTTTATGTT TATGTTAAG AGCACTTTG 1500  
40 TATTTTATT ATCAGTGG CCAATATT GTATGAGCA TATGTAGCAT TTCAAGCAT 1560  
GTTATCATG TAACTGCG GTAGAGCAA AGCTGTAAAG TAGATTATC ACACATGAC 1620  
45 TGCATACA CTTCAATAT GTCAATGTT TGTGTATGA ACTTAGAGC CAAAGCCAC 1680  
ACGAGGCG AGAATCCA ATTTACTCA TGTATCATC ATTATGATC TGTGTGTAG 1740  
AATATGAGG TGTAGGCTT CAGCTGCGA AGTTACATG AGAAGGCCA CACTGTGAA 1800  
50 GGTTTTATT TACAATCAC TTGATTAC ACACTAGGT AGAATATTT TATTTTACT 1860  
GTTTATACC CAGATTTAT TTCTAGTGT TTCTAGGCA AGAATATCA TAAAGTATC 1920  
55 CTTTCAAT GCCTTGAGA AGATAGAG AAAAAAGT TGTATATT TTAATAAT 1980  
GTTTAAAG TCAATTTCA AATGTCTGT ACCAGATGT TACTTGTCT TAACTGTTA 2040  
60 TATGACTTT CATGAGACT GAATAGCTT GCTATGACA CTTCTTAT CTTGAGGT 2100

TAATCTTTG CTTCACTTT CTACAGTATG ACATATGAT TTCTATCTT GTAAAACTTT 2160  
2196  
5 TATAAAT TTCTATATAA AATATTGAA ACTTAA  
(2) INFORMATION FOR SEQ ID NO: 283:  
10 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1185 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 283:  
GCAGTTAAG CTTCTGATA GGAAGAGAG TCTGACAGA GCACACAT CTGAGCTCC 60  
AGAGTGGG GATCAGCAT CAGATCCAT CTTGAATTC TCTAATAA CTTGTACTC 120  
ATATGATC TCACAAAGA TCTGTATTC ATCTGTGCT CCACTTCCC TCTGGGTCA 180  
GTAGATGTA AGCTGACCT TGGCAGGCT CTTAACATGA AGATCTAG CTAGACAGC 240  
25 AGACTCCCC ATTTATGGA AAGAAATTC AATTTATCT CTTATTATA ACAAATTTT  
TAAATGCTT TGGTTAAA AATCTAATG TCTGGGTGT GATCAGTCAG GAGCAGTAA 360  
CTATCACTCT TGGCATCTT TGTCACTCG GAGATCTTT GGGGCTGG AGTCTCTCT 420  
GTCCAGCT AAGGAAAAG CTTCAAGG GTAAAGCCA CAGAACCTC GGCAGAAAG 480  
GCTGTCAG GAGATGAAT GTTACAGAG GAAAGGAG GAAAGGGGT GGAACAGAG 540  
35 TAGAGCCA GGAAGGATG CCGCATGGA GACCGATGG GACACTCTA TTGTGCAAG 600  
GGAGGATCT TCTTCTTGA ATCTGACCA CAGCTAGTCT GAACCTTCT TGGAAATGC 660  
40 AGCTGTTTC CCAATGATG GGCACACTCT CCTCGAAG CAGCAATGT GCTCTTATC 720  
AGAGGAAA AGTATCCATC AGTGTGACA GAGGTCACT TCGACTTGC ATGACTCTT 780  
45 TCGCAGCCA CAAGAGTCC TGTAGAGT GAGGATCCC TAGTCTTAG CCTGTCCGTT 840  
TATAGATTA CGATTACAC ACTCTGCTA GTCTCTGAA TACAAACAG ATTGTAGTC 900  
CACTGTCCA TATGAGCTG TTGAGTTT TCCAGCTTA GTTCATGAC CAGCATGAA 960  
50 GGGAGTCT GAGGCAAG TACACAGCT TCAAGATGAT GTTCCCTTTG AACTATATGA 1020  
ATAGAGCTT GACCAAGCCA GCTCGAAG CAGAGTGT GAAGAACTG AGGAATATGA 1080  
TGATGATTT GAGAGGACT GCAATATCT GATGCACTG AGGAGAGT TTCTAGTTCC 1140  
55 TTTGATGAG AGCTGTTTC CTTCTCTAA GCAAGCACC TCCCA 1185  
60

## (2) INFORMATION FOR SEQ ID NO: 284:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1634 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

10 AAGGAAAGG GAGGTACG GAGGTACG AAGTACG CTAGGCTG AAGTTACG 60  
AAGAGCTG TGAGCTCC TGCTACAG AAGAGCTG GGTCTGCT GGTCTGCT  
15 AAGCTGGT TGACAAAG GAGGTGGA GTCTATTT GTACAGAG TACATCAC  
TTAGCTG GAGGTATG TGAGTATG CTAGTTGT TACAGTCT GGTCTGCT  
20 CAGAGAAA TTGTACAT AGCTTGAG TGATCACG GCTTCCCA GCTCCACG  
AAGAGCTG TGAGCTCC TGCTACAG CATTTACG GTTTTATG ATGAGCTG  
25 ATTCTGCT TTGTATTT AAGAGCTG ATGAGCTG ATTATATG ATCTGCTG  
ATACAGAT GATATATG GTGTGCTG GATATATG CTGATCTG GTTCCAGG  
30 GTTACAGT AAGATATG GATCTCTT TATATATG TTGATTTG AAGATATG  
GAAAGCTG AGTATATG GTTCTTTG GAGTATG AAGATTTG GATAGCTG  
35 TGATCTCT AAGTACAT TTGAGGCT GTCTGAGT TTAGGCTCT TTGTCTG  
TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
40 TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
45 TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
50 TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT TGATCTCT  
55 AAGATCTT AATATATG GAGGCTCT GAGGCTCT GAGGCTCT TGATCTCT  
AAGATCTT AAGATCTT GAGGCTCT GAGGCTCT GAGGCTCT TGATCTCT  
60 AAGATCTT AAGATCTT GAGGCTCT GAGGCTCT GAGGCTCT TGATCTCT 1500

TGATCTCT TGATCTCT AAGATCTT GAGGCTCT GAGGCTCT TGATCTCT 1560  
TGATCTCT TGATCTCT TGATCTCT AAGATCTT GAGGCTCT GAGGCTCT TGATCTCT 1620  
5 KTAGAGGCT GGTG 1634

## (2) INFORMATION FOR SEQ ID NO: 285:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1795 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

10 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 60  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
15 GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
20 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 180  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
25 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 240  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
30 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 300  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
35 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 360  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
40 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 420  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
45 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 480  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
50 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 540  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
55 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 600  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
60 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 660  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
65 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 720  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
70 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 780  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
75 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 840  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
80 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 900  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
85 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 960  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
90 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 1020  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
95 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 1080  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
100 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 1140  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
105 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 1200  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT  
110 TTTCCCTG GGTCTCTT TGATCTCT TTCTCTCT TGATCTCT ATGAGTAT 1260  
GAGGCTCT TTCTCTCT TGATCTCT GGTCTCTT TGATCTCT TGATCTCT



AAATGACCAC ATTGAAGGAG CACAATGCC TCTGTGTG ATGCCACTTC CCAAGGTGGA 1320  
GACAGTGGAA AGAAGCCGAG GACAGGAAG GATTGGGTAG GTGAAGGGGT CAGGGGACTG 1380  
GTAGTCACCC AATCTTGAG AGGTGCAMAA AGCACTGGGG GCTACCCGTT AGCTGCATCT 1440  
GCCCCGCTG TTTCGCCGTT CATGTACAA ACTGCCACTA CTATGTACCT GCAAGTGGGT 1500  
TCCAGAGATG GGGGAGACTC AGCTTCTACT CCCAGAGAGC TCCAGAGGCC CAGAGAGAG 1560  
AACTGCGCT CATTTCAGTC TGGTCTACAC CCACCTTCTG GTAGCCCTTC TCTTCTGT 1620  
AATTCGCT GTTTTTCAG ACTCAGCTCA AATAGTGCC CTTCTTTAGC CCATCCCTGG 1680  
CCCCAGCT GAGGTGATCT TTCCCTCTTC TGAATATTTA GAGGATTTAC TGTCTGTCA 1740  
GTTCGTTGG CAGGCACACA CAGTGGGATA AATTCATTTG TTTTGAATTC TGATT 1795

20

## (2) INFORMATION FOR SEQ ID NO: 286:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 858 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

TCTCTTTGG GTCTGTGGTG TACTCTGGG CGCTCTCTC CGGCTGTGG GTTTGGCCA 60  
GCTCTGGAG GAGATCTGG CTCAGTTTC GGAGCGGATG AATGCCCTGT TGTGCAATT 120  
TCTGAGGTG TTCCCGCTGA AGTATTTTGG CTACAGCCA GATCCCTGTA ACTACCAAT 180  
AGCTGTGGC TTCTGTGAC TCGTGGCTGG GTTCTGTCTG GTACATGGCC CACGATGCT 240  
GCAGAGATC ACTAATCTCT TCTTGATCT GCTCATGATG GGGGCTATCT TCACCTGGC 300  
AGCTCTGAA GAGTCACTAA GCACCTGTAT CCCAGCCATT GTCTGGCTGG GGTTCCTCT 360  
GCTCTGTAAT GTGGGCCAGC TCTTAGCCCA GACTAGAGAG GTGTGCAGAC CCAGTAGGA 420  
GAGACTCTA AGTACATTTCA AGGAATCTG GAAGTAGAGC ATCTCTGTCT CTTTAGCCA 480  
TCCAGCTGT ACAGCAGAA CATGTAGAA CACAGATCT ATCATCTGT TACCATATA 540  
ATATCCAGG TCAGCCAGTG TTGAAGAGA CATTGTGCT ACTTGGACT GCTTCTCTT 600  
TTTAGCTTTA CTACTCTTT GTGAGAGTA CATTGTATGC ATATTAAGT TCTCATGTC 660  
ATATGAATAT ACAAAATAG CAGAAAGAA ATTAAATCA ACCAAATTC TGATGCCCA 720  
AATAACCACT TTATATGCT TGGTATAGT ATACTCTGA ACTTTTCT GTGCCTTAA 780  
ACAGATATAT ATTTTTTTTT ATGAATAATA AAACATATA TCCATATTTA TTCTCTCTT 840

60

TTAAACCTT ATAACTA

858

5

## (2) INFORMATION FOR SEQ ID NO: 287:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

GAATTCGGCA CGAGCGGC CATGGGCTC CTGCTTTGG TGTCTGCTGT ACTGCTGGGC 60  
GGCTCTTGG CGCTGTGGG GTTGGCAAG CTCTCGGAG AGATCTGGC TCCAGTTTGG 120  
GAGCGATGA ATGCCCTGTT GGTGAGTTT GCTGAGTGT TCCGCTGAA GGTATTTGGC 180  
TACGAGCAG ATCCCTGAA CTACCAATA GCTGTGGCT TTCTGGAAT GCTGCTGGG 240  
TTGCTGCTGG TCATGGGCC ACCGATGCTG CAAGAGATCA GTAACTTGT CTTGATTTG 300  
CTCATGATG GGGCTATCTT CACCTTGGCA GCTCTGAAG AGTCACTAG CACCTGTATC 360  
CGAGCATCT TCGCTGGG GTTCTGTCTG CTGCTGATG TCGGCCAGCT CTTAGGCCAG 420  
ACTAAGAAG TGTGAGACC CACTAGAAG AAGACTCTAA GTACATTCAA GGAATCTCTG 480  
AAGTAGACA TCTCTGTCT TTTATGCCAT GCAGCTGTCA CAGCAGAAC ATGTTAGAAC 540  
ACAGAGCTA TCATCTGTT ACCAGTATA TATCCAGGT CAGCCAGCT TGAAGAGAC 600  
ATTGTGCTA CTTGGACTG CTTTCTCTT TTAGCTTTAC TACTCTTTG TGAGGAGTAC 660  
ATGTTATGCA TATTAACATT CCTCATGTA TATGAAATA CAAATTAAGC AGAAAGAAA 720  
TTTAAATCA CCAAAATCT GATGCCCA ATAAACCACT TTAATGCTT GGTGTAGTA 780  
TACCTCTGA CTTTCTTCTG TCCCTTAAA CAGATATATA TTTTITTTTA ATGAAATTA 840  
AACCATAT CCTATTTAT TTCTCTCTT TAAATCCTTA TAACTATTA HAAAAAAAA 900  
AAAAAAAA CTGA 915

50

## (2) INFORMATION FOR SEQ ID NO: 288:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1517 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

60

CCCTGTGAGCA ACTAGTGGGT CCCCCGGACT GAGAGAAATTC GGGCAGTGGT TCTAGATCTG 60  
AAGATCTCT GAGTCTCTCT GAGAGATCCA AAGGCTCCGG GAGCAGAGCC CCAAGCCCCA 120  
AAGAGAGCCC TCAGAGAGCC AAGAGAGAGCC CTCAGGTGAC CAGGGGTGAGC CCTCGAGAGA 180  
CGAGCTTAG GCTTCAGAGAG ACGAGAGAGC GCTTCAGAC GCTTAGAGCCG AACCGAGATTA 240  
CGAGCTTACT TGAAGCTTTC CCGAGAGAGC AGTGTGAGCA GACAGAGAGTG GCGAGAGAGTG 300  
GAGGCTGAGG TCGTTCAGAG CAGAGAGACA CCGAGGCAAT CTCATAGAGG CTCAGAGCCAC 360  
CTCCAGCCCT ACCGTGTGACT CAGAGAGACA GAGAGAGAGG TTCTGAGTCA AACGTGATTC 420  
CAGAGATGCG GCGTGTTCCT ATAGAGAGATA CTCTTTCAG GGGGCGGCCCA AGCTGTGACA 480  
AGTCAGAGAG TGAAGAGAGC TGTAGTTCAG CCGAGTCTTG GCGATCCCTCA CCTGATGAGG 540  
TTTCTGGTGT CAGCAGAGCA AATCAGAGTT CTGTGTGTTA CCGAGCTCTG GAGAGAGCTT 600  
TGAATCGGCT CTGAGATCCA GCGGAGAGCA TGTGCTGTGC AAGAGAGCTT GTGCTGAGAG 660  
TGGCTGTGCG GCTGTGAGAT GCGCTGAGAT TGTTCAGAGA GAGAGAGAT GTTACAGAGAA 720  
ATGTGAGAGC TGAAGATATC TTGTGTGATC CAGAGAGACA GATGTAGAGTG ACTTTGAGAG 780  
GCTAGAGCTT GCGATTCGCC TATTCGCCCA GTGAGAGACA CTTGAGCTTAC GTGAGAGACA 840  
GAGAGAGCT TCAGAGAGAG GAGCTTGAAT TTCAATAGCA TGAAGCTTCA CAGAGAGAGC 900  
GGGCGCTCCC GCGGAGTGA CTTCCAGAGC CTGGATTAT GCAATCTGAA GTGTGTCTAA 960  
GAGTTTCTGC CAGAGAGAAA TTGCTTCTCA AAGATAGAGCA GATCAGAGAG CAGAGAGAGCA 1020  
AGTTGCTTGC GAGTCAATTT TAAATAGAGC TGAAGTTTGT CAGTCAAGAA ACAAGCTTGC 1080  
GTCACTCTCA GAGATTTCGA GTTTTCAGAC TCCAGCCAGT TCGTCCGTGT GATTGAGCCA 1140  
GATATCTCG TTAATCAGAC GCGTTCGAGG GAGCAGCTCC CTGCTTTTGA GTCAAGTACA 1200  
GAGAGCTAC TTGAGCGGCC CCGTGGCCCC GAGATTCAGC TGAAGTTCG GATGTGAGCA 1260  
GTGAGCCCCC GTGAGAGACA GCAATGAGTC CTGAGTTCGA GCGTCTTTC TTGAGAGCTTA 1320  
CGAGTTTAAA TGTCTTCAAA ATGTTCAGAT ACCGCGCAAT GCGCTTCAGC AACCGAGAGC 1380  
TCCCTGAGCA GTGTGCTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GAGAGAGAGTG 1440  
GAGAGTACAC CTTTGGGTGT GCGGTGTGC CCGAGAGCT TGAAGTAAAT AATCTTTAT 1500  
TTCCAAAAAA AAAAAAA 1517

55

(2) INFORMATION FOR SEQ ID NO: 289:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3865 base pairs  
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

5 TGAAGAGAGG GAGCTTCTTT GAGCAGTGGG CCGAGGCTTG GCGCTTCACA CTTCATTTTC 60  
TGAAGCTTCT CTCTCTTCAT TTGATTCAGT GTCTTTCTTG CAGTGCCTTT TCAGAGAGAG 120  
TGGTTCACAT GGGGCGAGCT AACGTGAGT GAGAGAGTG GAGAGAGACA GATTCATTTT 180  
ACTCAGATCT TCTTAGTACA ATAGAGAGCA CCGAGTCTTT CTAGAGAGAG CTGGGTGTGTG 240  
GTCCCTTAGG TATGAGCTTC TGTAGATGTA CTCTCCAGAA ATGTAAAGCT TTCTATTTTC 300  
AGCTGTGAGC ACCGTGTGAG GCAAGTGGC TTGCGCATTG GCGCTGTGAG GTGCAGAGCA 360  
GCGTGGCTGC CCGCGAGAGC CAGGCTTCT TTCTGATGCC TCTTTCTTGA ACAGTGAAGT 420  
GAGCTTGAAT CTGAGAGAGA ACCTGTCTTT TTAGCTTCAGC AACAGAGAGAA GAGGTGAGCA 480  
TGAAGTCCCC GCGGCTTCAG CAGAGCTGGG AACAGAGAGG ATGTGTGAG AGCGAGTTC 540  
CTGTGGCTCT CTGCAGAGTG TTTTTCCACTA GTCACTAGTG TCTTTCTCTT GTAGCTTAATC 600  
AATCAATATT CTCTCTCTGC CTGTGGGAGC TTAGAGAGTG CTGTGGGTG TTAGCTTCAGC 660  
CTGCGAGAGC AGTGGAGGAA AAGAGATTAAT CAGTGAAGAC TGTTCATCTC AGAGCTCTTG 720  
ATCTAGCCCA CCGCTTAGAG TCGAGAGCTG GGTCAAGCT GATAGAGAGC AGCGCTTGGC 780  
AGCAGCTCT GAGATGAGTG GAGGTGAGAG AAGAGCTGAG TTCTCTTTGC CTCTCCCTTC 840  
TCCAGATTTA CTGAGACTCT ATCTGTTAG GATCTTTCGA GCTTGTTTTC CTGTGGGTG 900  
GAGAGAGAG CAGAGAGAAA GAGAGGTCTT AAGAGAGAGCA GCGCTTCTTT GTCTCTTGGG 960  
GTAAATGAGC TTAGACTTAGA GTAAATGAG AGACAGAGAG CCTGTGATTT TTAATTTCCA 1020  
TAAATGTGTA GAGTAAATTA TATAGATTA TATATTTCT TAAATTTTTC AGCTTTGAT 1080  
ATGTCTAAA ATCCATTCGC TGTGCTTGA AGCTTGAAGT AAGACATTA AAGAAAGCTT 1140  
GTTCATTTTA AAGATTTTAA TTAATATAT GAAAGTTTGC TGTGGCTTAT GCTTGTATAT 1200  
GTTCGGGGGA CAGGAGAGTG GTCTGGGCCC ACATTTAGAA GGGAAATAT TTGTGCTCT 1260  
GAGAGATTC GAGCAGAGTA TGGGCTCTTT CTGCTTAGTA CTGCGAGAGG GACTTTTAG 1320  
GTGTCTTCTC CAGTCTTAC CCGCAGCCC CCAATTAGCG TAAAGAGAGC CCGAGAGAG 1380  
GTGAGGGGCG AGCAGTGGC TCAGATTTG TTCTCTCTTG AGATGTACA GCTTCAGATCC 1440  
AAGACCTTG TTCAAGATAT TTAATTAAT TTAAGAGCT GATGGAGAT TGAAGTGAAT 1500  
GCGTTTAGA GAGAGAGCC ACCCTTAGG AAGAGAGAT CAGTCTTGC CATTCATCT 1560  
GCGTATGCT AACAGAGAGC AGCTGATTTA GAGATGTGCC CAGCTAGCT GATCAAGAGG 1620  
AATTTCAAG AGCGCTGGG CAGGCGCTTG TCCGAGAGC GAGCTTCAG AGTAAAGAGA 1680

55

60

5 CATTGGGAAA GTTCCCAACC ACTTGSTAGA CCACTAGGTT CTCTGTTTTC CTTTCCCTTT 1740  
CCTTTTCAAA TCCCAACAGT TTCTGTTGGG GAGAGCTGT AATTAGCTTA GTCCAGGTAC 1800  
CAGATCCAG CTAGGGGGC AGCTGCTTGG GATNACTCCA AGAAAACCTG GGCACACATA 1860  
TTTTTCCAAAT TATNAAGACT GTGGCATMAA TTTTAAATAG AGTTATATAG AAACACAGAT 1920  
TCTCCAGCTG CCAGAGGAG AGGTAGGGC TGGACTCCCT GCTGTGGGCC AGCCCTTGT 1980  
AGGGTTGGT CTCTCACTGC AGCCAGACAG GATGATCTTG GTTCTGTGGG AGGTTAGCT 2040  
GCCCCTTGCC GAGTCTCTGA CCGAATTAAG AGTCCAAACC CGCTGCTTCC GTGTCTTAG 2100  
AGATGGGTAA ATGGGTGATG GATGAGCAG ACTGAGAGCA CAGCAGATGA CTCATGTGTC 2160  
GAAGAGGGG GGAGATGCT GGGGTGGCTA GCTAATGTC CCGCTTTTCA CGGATTTACA 2220  
GGAAATGGAG CCGAGCTTGG TCAATAGTT GTTTTCTTTC CACTGTGCGA TGCATCTTC 2280  
AGAAATTTTG AAGTCAAGCT GCACCTTCTC GAAGACTTTC TTCTTGGGCT TBAAGCTTC 2340  
ATCTGTTGG CCGTTTTCAT AGCCCTTCAAC AAACACTGTC TCACCGAGAG CAGAGCTTC 2400  
CGAGGGTTC AGAGGTTTCA CTGGCGTTT ATCCCTTCTA TAGAAGCA CAAGAGATG 2460  
CCTTGGAGCT CGACTCTCT CATCTCTGG GGTTCAGGT TGCACAGAC CACTACAGC 2520  
CTCTCTGCA GTTCTCTTT GGGCAGCAG TGTACAGGC CGCTCACAC AGTCTGGT 2580  
TCAGTCTCC CGAGTCAAT CTCTCTTACA TACAGGCTGT CTGATCTGG GTCTTCTTC 2640  
ACAGTATGA TTTTCCCAAC ACGGATATCC AGCGGGATG GATGACCTC CTCTGTTCT 2700  
GAATCTTGG CAGGCTTTG GCAATGGCT TCTGCTTTCA GGGATCTGG TAGCCAGGC 2760  
TGGCAGTTT TTTCAAGGCA GGGGTATTA ACTTTTCCCG GATGATCC AGCAACTGT 2820  
TCAGTCCGAC TTCAACAGAA TTCTTCAGT CTCACAGATG TACAACCTCA GCAGCAAGT 2880  
CCTTTTCCAG GTCCAGTAA GCTGTGAGG TTTTGTTC ACCCAATTTC TCATCTGTA 2940  
GGATCAGAA CTGGACTTA AGGGAAAA GGACATCTTT GATGAGGAC AGAACCCAT 3000  
TGTCTTCCAC ATTTCCTGGC TCACAGAGG CCTTCTTCAG TTTTCTTC ACATCTCT 3060  
TCCATCAG GAGATCAATC TTGAGCTCT CTCTGAGCA GCTCAATTTG CTCCTGTTA 3120  
ATCTTGGAC CATGAGTTC ATCAGATGA CCGTTTTC ATAGCCAGT GCAGGAGGT 3180  
ACTTCTTGC AAGGTGAAA ATCTTCTCT GATCATGACC TCCAAATTTG GCATCTACT 3240  
TTAAATATC TTCAATCAAA GCGTCAATC CGGGTATTA GAGGCCACT AGCAAGGCT 3300  
GCTCCAGCTG CTTTACAGC TCAGTCCAG CTTTCTTGA ATCTGTGCT GTGACAGG 3360  
AGGAGGTCT GTACATCT ATGTGTACT CTTTCTGAG CTGTTATCA GTCCCTTGA 3420  
TGAAGTTGAG CTTTCTCAG GGCACAAA TGCTCTCAG CATTTCTTG ATCAATCT 3480

5 CATGTAACCT GACTCGAGT TCTAGAGTT CCAATGGGC TTTCATGTTA TCAGGTATG 3540  
CGTGAAGTTC CGCAACAGAA ATTGTTACT CACACCTGCG CTTTAAGAG TCTGCAATCT 3600  
TTGACATGGG CACAAAGTAA GGCACATGTG GTTTGCCCCG GTTGGCTTT CCCCAGTAA 3660  
TTTTAAATTC CCGTCTCTTC AGTATCTCT TCAAGCTTTC TTCCCCAGA ACCTCTTGA 3720  
GTTTCCGGT GATNAGTGC AGTTTCTCTT CAGGGCTGGG AGCTTCCCC ATGCTCCGT 3780  
ACCCCTGCT CCCCCGCTCA GCGCGCACC AGAGCCCCCT CTTGGGTCA CGTGCGCGC 3840  
GCTTCCGGG AACTGTACAG CGAGT 3865

(2) INFORMATION FOR SEQ ID NO: 290:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1910 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

AGGAGAGCA GAGAGGGG TCTGCGCGG GCGCTACCC AGAAGCAGC GAGCGGAGC 60  
AGGAGTGG CTGTCCCGA GCGCAGCCC GAGCGAGCCC CCCCCCGCC CCGGAGGAC 120  
GGCTTTCA GCGAGCCGA CTCTTAGAG GAGGGGAGGC GCGAAGCAG CTCAGGCTTC 180  
ACCCAGCCG CTGCCCCAG CCGCGCACT CCGAGCTCC TCGGACTCG CGCGTCTTC 240  
CTGGAATCT CGAGGGGAC CGACTGTGCA GAGCGCATG AGTTGGTGT GGTCTTCTTC 300  
TCAGGCTGC TGGCCCCCAT GGTCTTGGCC AGTGCAGCTG AAGAGGAGA GGAATGAGC 360  
CCTTTCTAT ATGATTACCA GACCTTAGG ATTGGGGAC TGGTGTGCG TGTGTTCTTC 420  
TTCTGGTTG GGATCTCTCT TATCTTAAGT CCGAGGTGCA AGTGCAGTTT CAATCAGAG 480  
CCCCGGGCC CAGAGATGA GGAAGCCAG GTGAGAGACC TCATCAGCC CAATGCAACA 540  
GAGCCCCA GAGCAGAAA CTCAGTGA GGCATCAGT GGAAGCTCTT GGAAGCTGAG 600  
CGCGCTCTT GAACCTTTGG ATCCAAATGT CGATGCTTAA GAAAGCCGC CACTTCAGCA 660  
ACAGCCCTTT CCGCAGAGA AGCAAGAAC TTGTGTGTC CCAACCTAT CCGCTCTTAC 720  
ACATCTCTC CAGCTGATGA TGCATCTTAC ACTTGCCTCC CCACTGAGC CTGCGGTCT 780  
GCCACCTCC GGTGATGTGT GTGTGTGTGT GACTGTGTGT GACTGTGTGT GTTGTCTTAC 840  
TGTGTCTTT GTGCTACTT GTTTGTGAT GGTATTTGTG TTGTAGTGA ACTGTGACT 900  
CGCTTTCCA GGAAGGGGT GAGCCAGAT GGCATCTCT CCTCCCTGCC CCGGTGGCCC 960

1020	TCGATCAAGCT TCTGCTCTTA GAGAGCTGCT TGTTCGCCGA GACCAAGCCC CTCCCTCAAT
1080	TTAGAGATCC GTAGAGTAGA AGCAGCGGCA GTGTGTTCA GTGCTTTGAG GACTCGGAA
1140	GGTTTCAGAC ACTTGTGTAI CATCTGTAT GAACTGCTTT GACTCTTTA ACAAATACCT
1200	TGCTTCCTTA TCCCACTGTA TCCCACTGTG AAGGTCTCTT AACACTGGA GATCAAAAC
1260	AAGAGCTGCG TGAAGCCAGC GTTGAAGTCA GCGAGCTAAT GCGCTTCCTT GATTATATTC
1320	TTCCCAAGGG CTTCACAGAG GAGTCCCAAT CTGCCCCCCC CATTACAGAA GCGCCCCGAG
1380	ATTCCAGGCC CAGGAGTTCT ACTTCAGCCC TGGAGAAAT GTCCCTCTCA TATCTGTCA
1440	GCATTAATAC CATGAGCTCT GAGAGCTTAC CCGTTCCATC CTTCCTGCTT TCTGAACCT
1500	CAATCTAGAG CCGAGCTCAAT CCAAGTCAG AGTCAAGTCC CTGCAATGCG GTCTCTGCGA
1560	GGCAATGATT GAAAGAGTCC TGTTCCTGCG GAGCAGCAGC ACCGAGATCG ATTGAGGAG
1620	ACAGAGGCC TTTCCTGCTTC TCCCTAGCTC CCGTTAGATG GCGAGCAGAG GCAACTCCCG
1680	CATCTTCGCG TCTGCTGCTC GATGTCTAGA GCGGTAGAGC AGGTGGTTG GAGAGTCAGC
1740	AGCGTCCGCG CAGCCCTTGG GAAAGATGAG AGGTGAGAG TCAATAGAG AGTGGAGACT
1800	CAACCAAGAT CCGGCCCCCT CTGTCTCTCT TGTTCGCCCG GAAACCAAC AACCTTCGCG
1860	CTGTAGCCCA TGTCTGTTCT CTGTATCTG AGTATCTTC AACAGACAA GAAATTAAGA
1910	ATTAATATAT CTGTCTTTCA TAAATATATA AATATATATA ACGGCGCGCG
35	(2) INFORMATION FOR SEQ ID NO: 291:
40	(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 3276 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
45	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 291:
50	GGCAAGCTCG TTGAGATGCT GGTGCGCGCT GCGGCGTACA CTGCACTGCG CAGCTCCGCG
120	ATTCAGAGCC AAGGTGTTTC GCGGCAAGCC TCCCTGCGCG TGGAGAGAG CAGAGAGGAA
180	GCGGCTTCCG GATTGCTGCG GCTTCTCTCG CCAAGCGCTG GAAAGAGCCA GCGCGCGGCT
240	AGTTCGCTCG TTTCACAGCC TGTAGAGCTT TAAAGCGCGG CCGGCAAGC CCGCGCGCGG
300	GAGAGGCGCG GCGCATGACT TGTAGAGCGG ATTCAAAAGG TGAATGACTA TCAACAGCA
360	TTCTCAACA GAGAGAGCT CCAATGCTT TAAATTTTGA TGAAGCAATC AATGAGAGCA
420	AAGATGCTGT GTCTCTTCTC CAGCCAGAGA TGAATGAATT GCAATTTGTC GAGGTGACA
480	CAGTGTCTCT GAAGGAGAG AGAGAGAGAG AAGCTGTTTG CATCTCTCTT TCTGATGATA

540	CTGTCTCTCA TGAAGAAATT CCAATGAATA GAGTGTTCG GAAATACCTT CGTGACGCC
600	TAGAGAAATG CAGTACAGAT CAGCCATGCC CTGATGTGAA GTAGCGGAAA CATTATCTATG
660	TGCTGCCCAT TGAATGACAA GTGAGAGCCA TTACTGTGTA TCTGTTCGAG GTATACCTTA
720	AGCGTACTTT CTTGAGAGCG TATTCAGCCA TCCGAGAGAG AATCAATTTT CTGTGCGCTG
780	GTGGAGATCG TCGTGTGAG TTCAATGTCG TGAATACAGA TCTTACCGCT TATTCGATG
840	TTCTCTGAGA CAGATGATCT CACTGCGAG GCGAGCTTAT CAAACGAGAG GATGAGAGAG
900	AGTCTTTGAA TGAATGAGAG TATGATGACA TTGATGTGCTC CAGAGAGAGC CTAGCTCTGAA
960	TAAAGAGAT GTTGAAGCTG CCGGTAGAGC ATCTGCGCT CTTTAGAGCA ATTGTGTGTA
1020	AGCGTCTGAG AAGAAATCTG CTTTAGAGAG CTCTGAGAG AGGAAAGAGC CTGATGTGCTC
1080	GAGCTGTAGC AATAGAGACT GAGCCTTTCT TCTTCTTAT CATGTGCTT GAAATCATGA
1140	GCATATGCG TGTGAGATCT GAGAGAGCC TTCTGTAAAC CTTTGAGAGAG GCTGAGAGGA
1200	ATGCTCTGCG CATCATCTTC ATTGATGAGC TGAATGCAAT CCGTCCGAAA AGAGAGAAA
1260	CTCATGAGCA GGTGAGAGCG GCGATTGTAT CAGAGTTTGT GAGCTGTGATG GATGCGCTTA
1320	AGCAGAGGCC AACATGATAT GTTATGCGAG CAGCAAGAG ACCGAGAGC ATTGAGCCAG
1380	CTCTAGCGCG ATTGTGTGCG TTGAGAGAGG AGGTGAATAT TGAATTTCT GATGTAGAG
1440	GAGCTTGAAG GATTTCTGAG ATTCATACCA AAGAACATGA GCTGCGAGAT CATGTGAGAGC
1500	TGCAACAGTA GCGAATGAGA CTGACCGGCA TGTGGTGTCT GACTTACAG CCGTGTGCTC
1560	AGAGCTGCTT CTGCAAGCCA TCCCGAAGAA GATGATCTTC ATTGACTTGG AGGATGAGAGC
1620	CATTAATGCC GAGGTCAAGA ACTCTCTAGC AGTTTCTATG GATGACTTCC GGTGCGCCTT
1680	GAGCAGAGAT AACCATAGAG CAGTCCGCGA AACGTGTGTA GAGCTGAGC AGGTAACTCG
1740	GGAAGAGTTC GCGGCGCTTGG AGGATGTGTA ACCTGAGACTA CAGAGACTGCG TCCGATATTC
1800	TGTGAGAGAC CCAAGCAATT TCTGAATATT TGGCATAGCA CTTTCAGAGG GATGTCTGTT
1860	CTATGAGCTT CGTGCGCTGCG GAAATACCTT GTTGGCGAAA GCGATTTGTA ATCAATGCA
1920	GCGCAACTTC ATTCAGATCA AAGGTCTCTGA GCTGTGAGC ATGTGCTTTG GAGAGTTGTA
1980	GCGCAATGTC AAGAAATATT TTGAGAGAG CCGCAGAGCT GCGGCGCTGCG TGCATTTCT
2040	TGATGAGCTG GATTCGATTT CAGAGCTCG TGAAGGTGAG ATTGAGATG GTGTGCGGCG
2100	TGTGTAGAGTA GTTATCTAGC AGATCTGAG AGGAAATGAT GCGATGTGTA GAAATATATA
2160	TGTGTGATCT ATTGGAGCTA CCAAGCGGCC TGAAGCAATT GATCTGCGA TCTCTGAGAC
2220	TGCGCGCTTT GATTCAGCTA TCTACATGCC ACTTCTGATAT GAGAGATGCC GTGTGCGCAT
2280	CGTCAAGGCT AACGTGCGCA AGTCCCGCAT TCCCAAGAGT GTGAGCTGCG AGTTCCTGCG

521

522

TAATAATCACT AATGGCTTCT CTGAGCTGA CCTGACAGAG AATTGGCCAC GTGCTTCGAA 2340  
GCTGGCCATC CCGTAATCCA TCGAGAGTGA GATTAGGCGA GACGAGAGA GCGAGACAAA 2400  
CCCATGACC ATCGAGGTAG AAGAGGATGA TCCAGTGCCT GAGATCCGTC GAGATCACTT 2460  
TCAGAGACC ATGCGCTTTC GCGCGCTTTC TGTCACTGAC AATGACATTC GGAAGTATCA 2520  
GATGTTTGGC CAGACCTTTC AGCAGAGTGC GCGCTTTGGC AGCTTCAGAT TCCCTTCAGG 2580  
GACAGAGGT GAGCTGGCC CAGTCAAGG CAGTGGAGGC GGCACAGGTG GCGATGTATA 2640  
CACAGAGAC AATGATGATG ACTGTATGCG CTAACTGTG GTGGCCAGCG TCGAGTGAGC 2700  
TGGCTTCCTT GAGCTTGTAT CCTGGCGGT GCGCGCGCTT GCGCGAGAGA GGCACGAGG 2760  
GTGGGCCAC AGCTGTCTCC ATCTCCAGT CTGACAGTT CAGCTACAGT CTGACTCTGG 2820  
ACAGGGGGTT TCTGTTCGAA AATATACAAA CAAAGCGCAT AATATTAAG CAGTTTTCAT 2880  
TTGGTAGCGG GAGAGTGAT TACCAACAGG GAATGGGCC TTGGGCTATG CCAATTTCTGT 2940  
TGTAGTTGG GCGAGTCCAG GCGACTGTG TGGGTTGTGA ACCAAGGCAC TACTGCCACC 3000  
TCCACAGTA AAGCATGCG ACTGACTCA ATGCTCCCG AGCCCTGCTT TCCCGCTATC 3060  
CAACCTGGGT AGTGGGTAG GCGCCACAGT TGCTGCATGT TTATATAGAG AGTAGGTGA 3120  
TTTATTTTAC ATGCTTTTGA GTTAATGTTG GAAACTAAT CACACAGAT TTCTAAGCA 3180  
AATAATGACA TGTCTTAAAA GACATATAA CTTTGGGTCT AATGGGTTA AAAAAAAAA 3240  
AAAAAAAAA GCGCTCTCTA AAGTCCANN CTTCGT 3276

(2) INFORMATION FOR SEQ ID NO: 292:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1695 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

TTGCCATGCT TGAATTCGCC TCTCAGGCC AGCTTAGAG AAGAGTTGG TAGTCCGAGA 60  
GCTGAGGCAG GAGCGGCGAG TTTCTGGCGG GTGAGGCGCG AGCTGAAGTG ACGGCGAGG 120  
CGGAAGCAC GTTGGTTGGG CCGAGAGAG GCGCTGGGCC CAGGAGGAGG AGGAACCTT 180  
TCCGAGAAA CAGCAGCAG CTGAGCTCT GTGACGAGG GGCACAGAT GCGCGCGCG 240  
AAGGAGCCT CTGGGTGAGG ACCCAACTGG GCGTCCCGCC GCTGTCTCTG CTGACCATGG 300  
CCTTGGCGCG AGGTTGCGGG ACCGCTTCGG CTGAAGCATT TCACTCGGTC TTGGGTGATA 360

GCGCTCTTTC CGACCGGGCC TTCTAGTTGA CTTACCCCTT GCACAGCTAC CTTAAGGAGG 420  
AGAGTTGTA CCGATGTCAG AGAGTTGCA GCGCTTTTTC AATTGTGTCAG TTGTGGATG 480  
ATGGAATGA CTTAAATGCA ACTAAATGG AATGTGAATC TGCATGTACA GAAGCATATT 540  
CCCAATGTA TGAAGATAT GCTTGCATC TTGTTGGCA GAATCAGCTG CCAATCGCTG 600  
ACTGAGCA AGACAGACTT ATGTCCCTGA TCCCAAAAT GCACCTACTC TTTCCTCTAA 660  
CTCTGTGAG GTCAATCTGG AGTACATGA TGGACTCCGC ACAGAGCTTC ATAACTCTT 720  
CATGCACTT TTATCTTCAA GCGGATGAG GAAATATAG TATATTCAG TCTAAGCCAG 780  
AATCCAGTA CCGACAGAT TTGAGCAGG AGCTTACAAA TTTCAGAGAA TCATCTCTAA 840  
GCAAAATGTC CTATCTGCAA ATGAGAAAT CACAAGCCA CAGGAATTTT CTTGAAGATG 900  
GAGAAATGA TGGCTTTTTA AGATGCCCT CTCTTAATC TGGGTGATT TTAACCTACA 960  
CTCTGTCTT CTGGTGATG GTATGCTTT GGAATTTGTT TGCACCTGTT GCTACAGCTG 1020  
TGCACAGTA TGTTCCTCT GAGAGCTGA GTATCTATCG TCACTTGGAG TTTATGAATG 1080  
ACAAAAGCT AAACAGATAT CCACTTCTT CTCTTGTGT TGTAGATCT AAAACTGAG 1140  
ATCATGAGA AGCAGGCCCT CTACTACAA AAGTGAACT TCTCATCTT GAAATTTTAC 1200  
CATTTTCTT TTAAGAGCA AGTGTATAG ACATCTNAAA TTCCACTCTT CATAGACTT 1260  
TTAAATGCT TTCAATGAT ATAGCCCTTA AGAAATCACT ATAAATGCA AATAAGCTTA 1320  
CTCAATCTG TGAAGACTGT AATGCTATA ACTTTATGG TATGTGTTTT GTAGTAATTT 1380  
AAGAGTGA TGTTTGGAT TGTATATTA TTTTACTAAT ATCTGTAGT AATTGTTTT 1440  
TTGCTTGGT TATGTTTTT TTCCCTTTC TTAGCTATGA GCTGATCAT GCTCCTCTC 1500  
ACCTCTGCC ATGATACTGT CAGTTACCTT AGTTAACAG CTGAATATT AGTAAATG 1560  
ATGCTTCTG TCAAGATGG CCGCAATC TGTATTTGA AATTAGCAG GAATGACCT 1620  
TTATGACAC TACATTTCA GGAAGTAAA TCATTAAAT TTTATTTGAA TAATTAAAA 1680  
AAAAAAAAA AATCT 1695

(2) INFORMATION FOR SEQ ID NO: 293:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1501 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

CACCTTCAG AGTCTTTTC TCTTTGCT TCTACTCAA ATAGCCGAG GAGTGGGCTT 60



5 GGAGCTTGCA TTTCAGTTTG TTGTGATGC CCGCGAAMAA GTGGGGSCA AACAGCTGA 1740  
AGATGCTGCC AAMAGCNAA CCATGGSCA GTGATGATGG AGCTAGGCC TCCAGCTCA 1800  
CAGCTCAAGA CGAAGATGAC GTTCTCATAG TTGATTTGGA TGAAGAGAT TCTTCAMATA 1860  
ATGCGAGCT CATGAGAGAG AGAGAGCCG CAGAGAGAAA TTATGATAGA AAGAGATCT 1920  
CACTGCAAG AGCTCAGTA TAGACAGAA GGAAGAGCTT GATGATGCA TACATPAGA 1980  
TTGACAGAA ATGCTCTTAA ACAGAGCCT CTTACTATT AGTTTATCTG GGCAGACCA 2040  
GATTGTATG TCTTTTGTTC CAAAGGGAAA AATTTGACAG CAGTACATCT AATATGATTC 2100  
TCTTCCCTTT GAAGCANTC ATTTTGTAG AACTGTAGA CACNTGCG TATGCTGTAT 2160  
TGAAGTAGG AATATGATT TAAAGCCT TTGAACAAAG TGTGTGCATA ACCATCATG 2220  
AGATAAACA ACAAMTGA TTTTGCCTT TTATGTAAA TACCTTAGG TATCATTAAT 2280  
AGTTTCAAAA TATTTGTGTT TATTAAGTT GATACCTGCT TATTAATATT ATGCTTTAT 2340  
TTTGTGCTAG AAGAGAAATT ATTTTAGCC TAGATCTAAC CATTTTCATA CTCTTAAGT 2400  
ATTGAACAG ATTCAAGAA GTATCGAGTG CTATGCAATG AACTTTTATT TTAAATGTTA 2460  
GATGGACTA TGTATATTA TGTAAACAA TGTATATTA CTCAGTTT CAGTTTGTAC 2520  
CGCTGCTAT GTCTGTGTA GAAGCCAA TTGTGTGATT GTTACAGATT CAGTTTATT 2580  
ATATTCGATG TTTTGTAAA CTCAAATAC GACTATATT ATGACCAAA TAAATGCAV 2640  
TGCNTTCTG TTAAMAAAN NACAGAAA AAAAAAACA AGA 2683

35

## (2) INFORMATION FOR SEQ ID NO: 295:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1454 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

50 GGACTCGGGG TGGCTCTTAG GGGCAGGGAT AGGGCTGGG AGCGCGGGC TGTGGCCGTG 60  
ACCAGCCCT TCTGTGCGAG GTTCCACCC GATCGAGGTG GTCAAGTCT TACCGGGGA 120  
CAGCTACCTG AGCCACTGCT TCTCCAGCA CTTCAATGTC GTCTGTCTCT CTCTGAAAG 180  
CAGCCCTTG CCGAGCCTG TTGAACAGGA CTTCCTACTCC GAGTTTGGGA ACAAGACAC 240  
AGGGAGATG GAGAACTAG ACCTGATCCA CTCTAGTGGC GTCAAGTTTA CTTACCCGAG 300  
TGAGGAGGAG ATTTGGGACC TGAAGTTTAC TGTGGCCAA AAGATGGCTG AGCCAGAGA 360  
GGCCCCAGCC CTCAGCATCC TCTGTAGCT GCAGGCTTC CAGGTGGCCA TGGCAGCCCC 420

60

5 TGGTCTGTC AGGGGCCCC TCGCCCCAA GACACTCTG CTCACAGCT CGAGATCTT 480  
CTCTCTGAT GAGGACTGTG TCCACTACCC ACTGCCCCGAG TTTCGCCAAG AGCGCGGCA 540  
GAGAGACAGG TACCGGCTGG AGATGGCCG CCGCGTCCGG GACCTGAGCC GAGTCTCAT 600  
GGCTACCCAG ACCTACCCGC AGCCCTCACC CTCTGTTTGG ATGAGGTGCA AGTCTATAC 660  
CTATGCGGCA GTGTCACTCT GAGCAGCTTT GGGAGAGTGC CAGGTGGCCC GECTAGAGCC 720  
AGCCAGGGCC GTGAGTTCCA GTGCAAGGTG TTGTGCCCCA GTGCTGAGAG CAGAGAGAAG 780  
CTCATCTGC TGTGGCTCG CCAGTGGAG GCGCTGTGTG GCTGAGCTG CCTGTGAGC 840  
TCACCGGCTA GCGCAGGCA CAGCAGCCT GTCTGTGTCA GCTGAGGCC TACTTGGGCA 900  
GGGAGCAGG CTTTTGTGTT CTCTAATAAT GTTTTATCTT CCGTTTGGTA CTTTAAATTG 960  
ACTGTCTCG CAGAAATGTG AACATGTGTG TGTGTGTGT TAATCTTTTC TCATGTGGG 1020  
AGTGAGAAAT CCGGGCCCC CAGGGCTGTT CCGTGTCTCG TCAGCTCCC ACAGGTGTAT 1080  
CAGCCTGCA CACCAATGTC GTGTCTCTCG TTGTGGGACC GTTGTAAACA GGTGACACTG 1140  
TGGTCTGAC TTTTCTCTCT ACAGTCTCTT TCCTGAAAGT TCGAGTCCAG TCTTTGTGTG 1200  
CTGTCTGCT TCTGTGTCT GTTCTGTGTG GCATCTGCT GTTATCTCG AGGCTGTAG 1260  
CAGATGATC ATTGAGACT CCGACCCAT ATTGTCTTC AAGTGGAGG TCTCCCTCA 1320  
TCCAGACAAG TGGAGAGCC CCGGGGCA GGGAGCCTGG AGCTCCAGC ACCAAGCTG 1380  
ATTCTCTG CCTGTATCT CTATTCGAT AAGCAGAT TTGACACCG AAAAAAAAAA 1440  
AAAAAAAAA AACN 1454

## (2) INFORMATION FOR SEQ ID NO: 296:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 828 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

50 ACCCTGGAT GCGCCCAAA CAGATCAACA GCGAGCTTAC ACAGGCATTA ACTCTCTCA 60  
ATGAGGAGA ATCATTTACA ACTGAGCAG ACATCTATAT GATCATTTAA GGAAGTGTTT 120  
CCCTTATGT TTAGCAAGTA TAATGGCTA ACTCTAATAT CCGAATGAAT AGTCTTAGGC 180  
TGGAGCAAA TGGGCTCAAA TTAGGAGAT AAGACATCA GTCCCAAGTA ATGATTCAT 240  
AGACTATCT AGCAGCACT ACCATAGCA CTATGTTAGG AGCTGAGG CCCCAAGTA 300

60

	GAAGATGTC ATTAATGTC CTCCTGTGA GGTCAAGAA GAATTCAGC ACGACACTA	360
	CAGTTAAC TGACTCAG CTCGAGTGA TAGCAGAAC ACTGAGAAA ATACTTATG	420
5	AGGGGGCAG GCTGAAGTC GGCCTGAG GATGATGGA ATTGATGAG AGATGAGAA	480
	AGACAGAGG ATCCAGATG AGAGAGCAT GAAAAATGAG CAGGGGCTG GATCAGTGG	540
10	GTATATCAG AGCACTTTC CAGATCAGC ATGATGCTC AAGTCCCTT GCTTATGCT	600
	GGCAGATGT CCGAGCAGA TGTGTGCCC GAGCCCATGT CCAATTCAT GTCTTCAT	660
	GGCAACTCA AAGGATCAT GTTCTGAAA GCTTCCCTC GGTATCAGAA ATCGAAATTA	720
15	ATCAGGAGTc TTTTCAGAT GGTGTTTTT GCTCTTGAT CTTCTATCA CTAAACTCA	780
	TCCTATCAG CTTTACAGAA TTAATTTTA TTTGTTTTG TGCATACA	828
20	(2) INFORMATION FOR SEQ ID NO: 297:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2416 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
25	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2416 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 297:	
	TCATTTTCA TTACTCAGA TCGACATTC TCAATTCAG TTTTCAGGC TCTCAGCTT	60
	AAACAAACCT ACTATACAA TGTATCTTCA ACGACAGAG TCTGAAATTA GCGACATTT	120
35	TTTTTTCTT GAGAAAGAG TCGACTGGG CAGCACTTT AGTCAGAGG GAGCTAGTG	180
	AAATCTAGC AATGAGATC ATCGATAGCA GCTTTCCTC AATATGTGA CTCCTCAGG	240
40	GCTAACTGC TCTTAGCTTA GAATTAATCT TTACAGAGAA TCTAGCAAGT AAGTGGGTTA	300
	ATCCATACA TCTGTAACT AGTTATATAG CTTCAGACA TCGAGAGAGC ATCGAAAGG	360
	GATGAGACA ACCCGAAGA TATGAGAAA GGGCATGATG AACCCCTTC CTCCTGGAG	420
45	GAGAGAGG CAGACAGAG GAGAGCTGG AAGGACCTTA GATGTTTAA GAGAGAGAG	480
	GGGAGCTTT GACAGCTCT TGCCTTTTC CAGTTTCAG CAGTTCCTG CTCCTGCAA	540
50	CTCTAGGTC AGTAACTTT GCGAATTCG AGATTTCCC CAGATATGA GAGAGAGAG	600
	GAATTTGGG GGTGGAGAG GGTGTGCTT GTTTTAAAG GATTAATTC CTCCTTCAG	660
	TGTTTAAAG TAGAGAAA AATATGAGG CAGGTGAGC ATCGAGCTA AATTTGTG	720
55	TTTATATTT GTGATGCTT TGAATATGT TACCATTTG TAGATACAC ACTGAACTT	780
	GAATCTGTG TCGAATAAC ACTAATTTT TTTGGAATG TTACTGTCA AAGGCTCTT	840
60	CCCTCCCTT CTTTCTCTA TGTACTCTT TCAATCTGC TTACTATGTC ACCGAGCAA	900

	TACCCATCA AAGACTAGG CATGAAAGG GGCCTTTCC AAGTACCTC TGGGTGCTT	960
	AAGCGAGCT GTGCTCTCTG GTTTATGAG TGTATATAG TCCCTGAGC CTTTCTTTC	1020
5	AATGAGCT AAGAGAGAG ACTGAGACA GTTATCGAT TCTGTATCA GATCAGTAG	1080
	GATGAGAGG CCGAGGAGT GAGGGGCTT CTGAAATTT GCAATGGCT GTAGTACTG	1140
10	AGTTCTTTC CATGTACAG AAGCTTACG CAGTTAGAT TTACAGAGG AAGGCTAGA	1200
	TCATTCAGC CATGTATGAG CTGTGTGCA GGAATGTGA ACCGAGAAA CTGTCTATCA	1260
	GGCAAACTC AAGCTTTCC TTTTAAAGG CAGCAGAGG AGGAGCTGG TCTCTCTGT	1320
15	CTGTACTTG TGTACATGC GGTGACTGA GAACTGCAT CAGACTGGG AGCAAAAAG	1380
	GAGAGGTAT TCTCTTTCT CTCGCCACC CTCGCCCTT GTTACAGAA CAGATTTCC	1440
20	AGGGGTACA TGAATTTCTG AATTTTAA AATGTTTTT GATTGATTT TTCTGGAGC	1500
	TGATAGTGC TTTTACGAT GTTCATACC CTGCAAGCT CCGAGCTTAA TCAATTTCT	1560
	GTATTTTCT GTTACAGTA TTTGTGTGT TGTGTTTTT GCGAGCTAT TTTGCTCTA	1620
25	TTATATATG AGTATGAT TGTATCTCT TTTTCCCTA GGAATATGA TTTGTTTTCT	1680
	TGTATATAT TCTGCTTGT AATACCTGA GCAAACTGG GGTGTAGAG GTTATGTAG	1740
30	GCTGCAAA CAGAGAGAA GCGGTGAG TGTACTTGT CTCGAGAGC TGAATTAAT	1800
	GAATCTGTA GCTTTGAT CCAATCTTT GCAAGCTCA AATGTGACA GAACTTTCC	1860
	TCTCTCCC ACTGCCATG GAGAGAGC GAGCATCCCT AATTCAGAA TGTCTTCTCT	1920
35	CGAGCCCTC CCAATGCAT GCGAAAGAG GTTACTTTG GGAATGGGG GATTACATG	1980
	GATCTGTG TATGAGCA CTAGACTTC TCTCTCCCT CCGTCTCC CCGAGATCA	2040
40	CTTCTAGGA GAGCGAGCT GCTTGCAGG GATCTGTGT CCGTCTAAC TCGAGAGAG	2100
	CATCCAGAG CTTTGTACA GTCTTAAT CCGTCCCTT CTGTATAGA ATCAATATG	2160
	ATATGACTT TCGAGCTTA CAGTATCAT TGGGTATCT GTATATATG CAGTACTG	2220
45	GAATGTAA GGTGAGATA GTTAACTTT TACCTTTAT TCTATTTTT TTGTATATC	2280
	TCTCATGTA TATTAATAT TATATATGT GATGCTTTT AATACTTA ACCGAGAGG	2340
50	GAGCACTGC TCGAGCTTG CAGTGTGA TGAAGCATG TTTTAAATA AAGAGGAAA	2400
	ACGCAAAA AAAAAA	2416
55	(2) INFORMATION FOR SEQ ID NO: 298:	
	(1) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 545 base pairs	
60		



(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

5 GAATTCGGCA CGAGCCATCC YTGCCCTCTC CTTCATCTCT ACATGCACTT TGTTCGCTGT 60  
10 TTCTGACTCA GCACTGACTT GATTTGTGGC CAAGGATGTA CTTATTTCTT CTCAGAGGG 120  
CAAAAATGTG GAATAGTGTG TGTCCATGCC TCTCTCTCATG GCTTACCAC CTCGCCATCG 180  
TGTTTATCA GTAAACACCA GAGAGAGGC TGCTGGAACT GACTCTCTGG AACTCCCTGG 240  
15 ATGGTTTGGT GCAGGAATGT AGTAGGCATTA CAGTGTGTG GTTGATCTG GGCCTTCTGT 300  
ATGTGAGTAG AGAGTAAAA GGCACCCTTC TCTTTGACCT YTGCGGAAT CATCTCAAA 360  
GAGATGTTT CCAAGATGCT TCTGAAGATT GCTTAAATAT AGCCGGTTTC CACCCCTGTG 420  
20 AATGCATCCA TTCTAGATAG CTCCTTCACC AGGACAGAG AACTGATTTA CAGAGTGAC 480  
ATGAATACAT TCCATCCAG AATTTCGAT ACTTCATTT NAAATTTCTAC CTATTAAAA 540  
25 NAAAA 545

(2) INFORMATION FOR SEQ ID NO: 299:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1530 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

40 GGCCTCTGCTG GGCATCATAC TTGTACTCG GTAAACAGTT TGCCCACTTA CCGCAGATGA 60  
AGCTGCTTC CAGGCTCTC CCGCTCTGT AGTTTGGAG CAGGCATCT TCACAGAGGC 120  
45 TGTGTGCTGG CAGGAGATGT GTGGGGCCCC GCGCAGGCTG CTCGCTCCC GTTCAGTGG 180  
TTGGGCCCC GCGTGATCTC CCACCTGTG GAGCTGCAT TACTGGAGG ATCATCGGG 240  
CAGATGATGC CAACTGCGCC GGCATGTCTC AGCGGGGAG CATCTCTGAG ATGATCGAG 300  
50 AGCGAGGGGC CATCATGAG ACCCGGCATT GCACAGCCA GAAGGGGGAG GCTGTGTGG 360  
CGCGCTTGGC TGTGTGGAG CGCAGCGCTT TCTGTCTCC CATGTGCATC GTTCAGGTGG 420  
CGCATGTGAG CCGCGAGATC ACTTACACTT CCAAGCACTC TGTGGAGGTG CAGGTCAAG 480  
55 TGATGTCCGA AATACATCTC AAGGTGCCA AATAGCTGAC CAATAGGGCC ACGCTGTGGT 540  
ATGTGCCCCT GTGCTCGAAG ATGTGGACA AGTCTCTGGA GGTGGCTCT GTTGTGTTT 600  
60 CCGCGCANGA CGAGGAGGAG GAGGGCCGGA AGCGGTATGA AGCCAGAGAG CTGCAAGCCA 660

5 TGGAGACCAA GTGAGAGMAC GGGGACATCG TCCAGCAGT CTTCAACCCA GAGCCGAACA 720  
CTGTCACTTA CAGCAGATCC AGCTGTATCC ACTTGTGGG GCTTTCAGAC TGCACCTGTC 780  
AGCGTTTGT CCACCGAGGT GTACCATGTA AGCTCATGGA TGAAGTCCC GCGATCTGTG 840  
CTGCAACCCA CTGAGAGACC AATATGTGA CAGCTTCCGT GAGCGCATTT AATTTCATG 900  
10 ACAGATCAG AATAGGCTTC GTCATACCA TCTCGGAGC CATGACCTTC ACGACCAATA 960  
AGTCCATGGA GATCGAGGTG TTGTTGAGC CCGACCTGT TGTGCAAGC TCTCAGAGC 1020  
GCTACCGGGC CGCAGTGGC TTCTTACCT ACGTGTGGT GAGCCAGAA GCGAGGTGTC 1080  
15 TGCCTGTCC CAGCTGTGT CCGAGACCG AGGACGAGAA GAGCGCTTT GAGGAGGCA 1140  
AAGGCGGTA CTTGCAAGT AGCGGAAGC GACAGGGCCA CCGGAGCCT CAGCCCTAGA 1200  
CTCCCTCTC CTCACACTG TGCCTGAGT AGCATGGA ACAGGCCCC TGTCAATGA 1260  
CTTAGAGCTT CCCCCCTGG CCAAAAGCC AATTTCATTT GAGAGTGTG GTTGTCTGAA 1320  
20 GTTTTCTGAT CACAGTGTTA ACTGTACTC TCTCTGCAA ACTTACATAC CAAGCTTTA 1380  
TTTATATCAT TCAATATCA ATCTACACA GTGTGTCCC GAGCGCGGG AGGCTTGGG 1440  
CAGAAACCTT CCGGATGCT TCGAGACAG CTGTAGGGA TGGGAGAAC CCGACACAC 1500  
30 TATATAGCT GATGCTTGGC TGGGAGAA 1530

(2) INFORMATION FOR SEQ ID NO: 300:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 997 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

45 AGGTAGTGA AGACACATTA CACTTACCA ACAGAGAA GATTCCTCC CTTTATATT 60  
TAACTATGTT TACAGGGAAT GCTTACATG TGGCTTCCG AGNATTTGT CCAATATGTT 120  
TTGAGAGAC CTAAATGCCA ACACTGATT GAATGGTAA AAGACATTA TAGCCAGAT 180  
50 GAACACCTT GGGCACCCT TCAGGTGCA CCGTGGATGC CTGGCTCTGT TCCCAACAC 240  
CCCAATGAG ACATCTTCAG ACATGACTTC TATTTGCCAG CTGTCTAGT GCGAGGTCA 300  
55 TGAAGGAGAC ATCATTAAGG GTCTCTTTA TGCCTCCCTG CTGGAATCC ACCAGCGGC 360  
TATCTGGTT TATGGGGCTG GGGACTTGA TTGATGCTT CAATACATC ACTGTGTGG 420  
60 CACAGCTT GAGCCAAAGG TAGATGATA TGTCTCTCAG TCGTTAGAG AATACCTAG 480

540  
540  
600  
660  
720  
780  
840  
900  
960  
997

TTTAAAGCC ATCTATGGA CTGAACTTGG AAGACACTA TGAAGAGCTT GCTACCTGCG  
GGGAGAGAC ATGTGAGAAC ATGCTGAGAA CTTCCTGGGA CAGTGTGGT GGGAGAGCG  
GGCTTGGAA TTGCTGGCAG CCTTAAAGAT AAGAGAGCTG KATTTAAATT GTGGTAAGT  
AAGTCTTGG CTTTGGAAAT TGTGCTGGCG GTGATCTGCG CTGTGTCTGT GACCCCTAAC  
GCTTAAGCT CTTCACTTAA CTTCCTCACT AAGTAGAAT GAGAACTGCT GTGATAGGA  
GATGAGAGA GGGATATGCG GTAGAGCACT TGAATTCAGT TGAATGCTCG CTGATAGCTT  
TTCAATCTCG TGAAGCTGCG GTTCTTAATA ATTCAGAGCTT TGTAAAGCTG GAGAGAACT  
TTGATGAAA GAGAACTTTC CTTCTCTTAC TGTAACTTAA AAAATTAATA GCTCTGATTT  
CAAAATTAAG AAAAATAAAA AAAAATAAAA AACTGGA

20 (2) INFORMATION FOR SEQ ID NO: 301:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2345 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (2) INFORMATION FOR SEQ ID NO: 301:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2345 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900

TTGAGCCGGA GCGTGAAGGCG CCGAAGAGAA ACTGAGAGCG GAAAGTAGCC GGGAGCCGAG  
CAATTCAGAT CTGCTGCGGTA GACTGCTGCG ACGACACACA TGTGGGCTGCG AAGGCTGGTG  
TGTCTCGGGA CACTACCTTC TAGGGTTTTC CACCCAGCTT TCAAGCAAGCG CTCCCTCTGTT  
GTGAAGAAAT CGATGAGAGA GAATCAATGCG CTGTAAACAC CTAGAGAGGA KATTCGACAC  
AAAAAGAGA TTGGATGCGG GCGTGGAGAA ACTGCGCAGG AACCTAAGAA GCGAGCAATG  
GAAAGATGGA TGGAAAAAAT ATTAAAAATT GATCAAGATCG GAAAGATGCT TGTGCTGGGA  
GGGGCTGCTG TTGCTCTTGG ACGATTTGCG TACATAGCTT TGGAGCTGTC TAAATGAAAT  
GAGCACTTGG AAAAGCTGCT AATTTGCGCT CAGATATGCA AAGATTAAGT TCAATTCAGC  
TATTAATGACT TAGAGAGAG TATTTGTTTA AAGACTTTGT CTGCAATAGC AATGAGAGAA  
ACGCTGTGTC TCAATGAATTT CATGATAGGA GCGCTTGGCG TGAAGATGCG TGTAACTTTT  
GGAGCAATGCG TTAGAGCTGCG AATGCTGCTA CGATGACATAC CATATGAGCA GAGCCGAGCG  
CGAAAGAGATC TTGCTTGGTGT GCTACATCTT GGTGTGATGCG GTGAGATGAT GCGTCTCTCG  
ACATATTAAG GGGGTCTCTT TCTCATAGAA GCTGAGTGT ACGAGCTGCG CATGTGGGAA  
GCGCTCTGCA CTGTGAGCAT GTGTGCGCC ACGTAAAAAT TTCTGAACAT GGGTGCAGCC  
CTGAGAGTGG GCGTGGCTCT GGTCTTGGCG TGTCTATGCG GATCTAATGT TCTTCACTCT

960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1500  
1560  
1620  
1680  
1740  
1800  
1860  
1920  
1980  
2040  
2100  
2160  
2220  
2280  
2340  
2345

ACGAGCTGCG CTGTGCGCAC TCTTTACTCA GTGGCAATGT ACGGTGAATC AGTGTCTTTC  
AAGATGTTCC TTCTGTATGTA TAGCCGAGAA GTAACTAGAC GTGAGAGAT ATGACAGATG  
TAGAGAGTTC AAAAATATGTA TCCCAATAGC TCGATGCTGGA GTATCTAGAT GATATCAATTA  
AATATATTA TCGAGATGCG AACATGCGG GCACTGCGAG GCGAGAGAAA GAAATGAAGT  
GATCAAGCTT CTGCGTCTGCG TCGACATGCA AATATCTGAT TTAAATGCGCG AAGATTCAGT  
TAAATGATTT GTAGAGAGCG CTTCCTGTGA AGTTTAGAGG ATTAAGAAACA TGTCAATGTA  
TTAAATGATTT CCGATTAATGT GATGCTCAG GTCTGCTCTT TTTCTGCGAG AATTAATGCA  
GTAACTCTCT CCGAATTAAG GAGCAAGATT TTCAATCTCTC AGTTTGAAT GATTTTAAAA  
TGTTTGGTGT AATGTGAAAA CTAAAGTTTGG TGTCAATGAA ATGTAACTCT TTTTCTACT  
TGTAAATTTA GTAGCTGCG TGAATTACTA AAATTTGCA AACCTGTGTT TCGATATTTT  
TTTGAGCTGCG AAGATATGTT AATTAATGTC AATAGAGATTT TGGAGCTTGG GTAAAGGAGAC  
CGAGAGAGG GATGACAGCG CAGTCTTGG TTTTATTAAT TACTTAAGAC TTAGCACTTGG  
TGTATTAAT TATGTAGAGG CCGATTAAGA AAGTGTGGT ATTTGAGAAC AAGTGTGAT  
TGTAACTATC ATGCTGCGAA GTTAAGAAA CTGTCAATCC TGAACAGAGC AAGATGATG  
CAATCTGCGG CTGTGCTGTC TGAATGCTCT CTTCGATATA TGAATGTGAT CAATTTGAC  
TTGACAGAA TGTAAATCAT ACGAGAGATC CTGATGAGAA TTAAATATGT GTTATTTACT  
TTGAAATGT ACGAAAGAA ATTACTTTAA AACATCTCTC AAGAGAAAT ATTCAGAGCA  
TGAATTAATG TCGTATTTCC AAGATTAAGA CAGTAACTC ATGATGCTA AGTGTATTTT  
TATTTTGA TATTAATGA ACTGCTAAT TGAATAGAGC TTGCTCTTGT GACCTCTGCA  
AAGTTTGAAG CTTTAAAGAA AAGCTCTT TGTGCTTAC ACTGAAATTT ATGAAGAGAG  
TTTTCTCTC AAGACTTGG GTTCTGCGGA TTGCTCTGCA GACTAGAGAC TAAAGAGCAA  
ACGAAAGAG AACATGATCT GTCTTAATGA AATATATGCA CCGAAATGCG TAAATGAGAA  
AATGCTCAT TAAATGCGCG TAGAGAGCTT TTAATCTCT TACAGCTGTA GACCTTACT  
TTCTGAGAG AATTTAATCT CTTTGAATAC AAGAGATTA TATTTAGAGG GACTTAATG

50 (2) INFORMATION FOR SEQ ID NO: 302:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2369 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60

533

534

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

5 TTTTITTTTT TTTTNCAG ATCAATGTTT ATTATTAAT TCAGTAAA 60  
AGATAGATA CATATAGGG AATCCCTTAA AATTCAGTC TAGAGTTATA CACATCTAG 120  
TACTTTTGA ATGATGTTA ACACACAAA AAAAAATCT TAAACACCTG AAGCCCCAC 180  
TATTACATG GACTATGTA ATAAAAANT TTGCATTTA ATTGTGTTCA CATATGAT 240  
TTACATTATG AAACCAATGG TGATATACA ATAAAGTAT AAAGAAATAG TAAATATAA 300  
CTTTAAAAAG CAAAGTTTAA TAGTCTGACA ATCTTAATTA TCTTAATGT ATATAAAAA 360  
TTAAACATA GAGCTTCTG TTACAAANT CTTAATCTC TGGTTGTAA TCATTAATG 420  
CTACAAAT ACATCCACA TCTCTAGGA CTGCAATTTG ATTTTITTC CCAAGATGT 480  
GTGATGAT AAATGCAAT TGAAGCAGA TATTAAATTA CTGTGACA GAANAAGAAA 540  
CTCAAGTTG GTACTGTCA CAGGCTCTT CCCATAGAA ATTATATAA CAGTAGATA 600  
25 AATTTAAAA AAMTCTTAA AGGGGATGC ATAGCAAG AGTACCATA ATGCCACG 660  
TCAAAAATC CCAGACCAA TGAACACAC ATCTTTCTC TCTCTTCAG GACAGAGG 720  
TGGATTTTC CATCAATTA CAGATGTA AGCAGAGAG GGCACACAG TGTACACCTG 780  
ATTAGATCTT GCAAAATCT AGATGGAG CAGGGTGGC CAGAGAGG GATAATTTAT 840  
ATATATTTCA AACTATATC ACATAAATG GAATCAGCC CATCCCAAC TGGCTCTGTG 900  
35 AAACATTTGG ACCTTTAG TTAATATAT ACAGTGTA ATATACAT AGATTTGAT 960  
GGGACCAA ATCCAGGGA CATTTATAT TAGTATTTA CTGTCTGTT TCAATTTAA 1020  
AATATTTTG CTAAATATC ATCTCACTG AGTCTATAT AAAAAATCT CTATAGATA 1080  
CAGATTTTAA CCTTTGGTA GTTGAAGCC TTTTGTGAC TTCTGTCTGA ACTGTAGGCA 1140  
GAATCTAGA TGTACATGA CATATGGGA AACTCAGCT GAGTCTATC AAAGCTGTG 1200  
CCTATAGGA GCTCGAGGT ACTTTGAAG TCANAATGA CCAGAAACC AAACAGGTA 1260  
ACAGTAGGA TGGCAGAGG GAATGGAATG CCAATATGG ACTATATG ACATATAA 1320  
CAGAAAGG AGGCTCTC GTACACAG ANTCTGTAG ACCTACAAA AAAAAAGC 1380  
CAGCACTT TTTTGAAC AGAGCCAT TATAGTGTG GAATGATA ATTAAGAAA 1440  
ACCAAGATC ACAGAGAA AACTACTGG TTTACTTGA AAAAAAGGA ATGTTTAC 1500  
55 CCGAGCAGG TTAATGTTG AACCCGCA CACCGCCA CAGACTCA TTGGTGTGG 1560  
CCTTCAGCA TTCACTTCA GGTCTAGT CAGAAATG CCGACTCT TTGTAGCA 1620  
AATCTACTG CTGTCCAGA AGAGGCAA AACATCTC CAGGATCTC GAGGATAG 1680  
60

(2) INFORMATION FOR SEQ ID NO: 303:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1181 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

GGAGCTGTG GTTTCAGCTC GTGCGCTCC CCTGGGTTT GCGACGTTTA GCGACTATG 60  
CGCTTCGCC AGCGCGGCTG CAGACTGGG GCGGTGGTG CTGGTCCGG GTGATCTAG 120  
GCGCTCCCTT GGGCTCCAG CTGTTCGGG GTCTAGTGG GAGTCAAGCA CGGTTCGGG 180  
CCCGAGTGT CCGGAAAGT GCGGCACATG GCGCGCAGG GAGAGCATGG CTCAGCGAT 240  
GGTCTGGTG GACTGGAGA TCAAGGATT GGCATTTGAG AAGGACAGA TTAATGAT 300  
GCGCTGTG ATACTGACT CTGATCTCAA CATTTGGCT GAGGTCTTA ACCTGATAT 360  
AANAACCA GATGATTC TGCACGAT GTCAATGG TOTTAGGAGC ATCAAGGAA 420  
GTCTGGCTT ACCAGGAG TGAAGGAG TACATTTACA TTGAGCAGG CAGATATGA 480  
ATTCTGTC TTGTACGAC AGCAGCTCC TCGGGCTC TGTGCTCTG CAGAAATTC 540  
AGTTCTGA GATAGAGT TCTTTGAAA ATACATGCC CAGTTCATGA AACATCTCA 600  
TTATAGATA ATTGATGTA GCACTGTAA AGAAGTGTG AGAGCTGT ATCCAGAGA 660  
ATATGAATT GCACAAAGA AGGCTGCTT TATAGGGA CTGTATGACA TTATGAAAG 720

[illegible]

20

(2) INFORMATION FOR SEQ ID NO: 304:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1537 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

30	CTTTTGTGT TCCGACCAAT CCCACCTCTC CTGACCCCTG GACCTCTTAC TTCCCGAAGC	60
	CCACATCTTG CCACATCCGC GCGCGGGAGCT AACGAGGGTT TCGACGACGC GAGCCCTCVA	120
35	GGGACATGCG AACTCACGCG GCGCCGCGCG GCGCGCCCGC AATATGAGC TGCCCGGAA	180
	TGGGAGGGT TCGAAGAAA CATCCGAGCG GAGGCTCGAG CTGTGATGCA GATGGAGAAC	240
40	ATGCAATGAT CCCTGAGGTC TACCTTTCAG GATTTGGTTC AACCTGCATCA GCGCCTGGGC	300
	GAGGAGAGAG TTAGACGCTCA TCGACCTGAT GCACTCTGTC CTGAGAGAGA GATATGAGAG	360
	TTCTCTGGGCA TGAAGGCTTT TAAAGGACAG CTGACCCGCG AGCTGGCAGA TCGATATGAG	420
45	CAGGCTGGGA AAAAGACAGG CTCCAGGGCC TTTCAGCTTT AACGCACAT CCACATGCTC	480
	AAGACCTACT TTGATGTGGA GCTGCTCAGC GTGCGAGAGC GGCCTCTGGA GTCTGATGATC	540
50	CTTATCAGAG TGTCTCAACT CCCCCAGAA ATTGCAAGTC AACCTTATGAG ACCCTCTATG	600
	CTGGCTTTCA CTCTGGTTCG TATCTTACT CATGGATGCA AGAGCTGTCA GACTATTTATC	660
55	CGGAGAGGCA CCGTGAATGAG GACAGCAATT GGCACAGCTT TCGGCTACTG GCTTGGAGATC	720
	TGATCTTTCA TTATCTTCTT TGCCTTACTG TCGAAGGCC AGATGACATC GCTTGGAGATG	780
	TTGCGACTGC TGGGCTATGAG CCGCTTTGGG GATTTGCTAT TCCCTTTTCAI GACTGATGAT	840
60	ATTCACCTGC AGCCCTCTTT GTACCTCTTC TGGCTGTGAG TGGATGACATC GTTCACACTG	900

5	GGGATGTTGAG GAGATCTGGA GAGACTGAG GAGCCCAAG TCCCGCCAT CGAGAGGTC CCGAGAGCA TCCCTGCAT GCTCCCTGCT GCTGGGCTG CGAGAGCT CCTGAGAGCT AGAGCCAG CTGTGAGCT GAGCCTGAG TCGACTGAC CCGACTGAA ATTCTGTGAC	1080 1140 1200
10	AGTCTCTTT CCCGAGCTG CAGAGAGAG GAGAGCTTT AAAAGAGCT CCTGATGACA TGTTCCTGAG ATGGGTTTGG CAGTGCAG TGAAGCTGAG CTGGTAACT AGCTCTTGA	1260 1320
15	AGCTGTGAC ACTTCTGAA GCGAGAGGCT AAGAGCTCT GGCAGAGCT GCGAGGCTT GGAGCAGTG CAGAAATGG GTGAGCTCT TTGAGAGCC CTGCCAGCT ACCCTTCTCT	1380 1440
20	TCTCTTTAT CTGTCCACA TTGTCTGCT AAATTATGAC TTGTGATTGA AAAAAAAAAA AAAAAAAA AAAAAAAAA AAAAAAAAA GAGCCCT	1500 1537

25

(2) INFORMATION FOR SEQ ID NO: 305:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1493 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

35	TCGATCCGA AACCAATCCC TCGCAACGA AATCTTAGAC ATCCCAATAT AATATCTTAG	60
	TTTATTTCTT ATTCAAGATCA TTATTTGAAA TACCGAGCTC AGTGCCTGCG TTATATTAATNG	120
40	TTTAAATCCC TTACCTAGTC TTGCTCTAAT TTTTATATTT AAATGAGAT GAGCAAAAATA	180
	ACACATCAAT GCGTGAAGCA ATTATTTGGA CATTCTCTGT TACCAAAA GACTAATATCA	240
	GGAATGCTCT GAGCTGTCA ACAAAGCTGT AATATTAACAG ACATGTAAC TCTTTTGCGAGA	300
45	GCTGCAATTT AAAAGGAAA AGAGAGAGG CAAAGAGATG CAGTAAAGGA AACGACAAAG	360
	GGAAGAGAG ATTGAAGCTC AAGAAAAGC GAAACGCGAA ACAGATGCGC AAAAAAACTT	420
50	TGAGCAATCT CGAGATGCTC GTCTGACAG CTGCGAAAC TTCCAGCGA ATTGAGAGGG	480
	GAGGAAAGAG AAGAAAATTC GAGCTTCTCT GAGACACAGC AAATGTAAAA TCGAGCAACG	540
	TGAGTAGACCG CCCAAATGCA CAGGACAGGA ACCTTTCCCC TCGATATCTC CTCTCTGCTTT	600
55	GAGAGAGATC AATCTTCTCT CCCAGTTCGA CCCCAAGATA GAGTACATAT TCGCTTTTAG	660
	TGCATTTTGT TTTCAATAGG AATTAAATTC GATGACAGTA ATTCTTTTGT ACATTGAAT	720
60	GAGGCGCTTG GTTTAAAAA AACCTTCTCC CTCTCCCTCC CCCATGAGCA ACCGATATTA	780

5 GAGGTGCCA CCATGTGTGC TGGCTCTCTCT TCCACAGGC TGTACTCTAG TGTTTTGTAC 840  
TTCACTGAAT TGTGATGGTT AGAAACTTGC TGTATAGTTT GTGGAAATCA TCCATATAAA 900  
CATACTGCTT AAACAGTGT TGTGTGTACT TCAGAGCAA GCTGTGAGG GGCACCTTAG 960  
GAGCCCTT GCTTCAAGTT GCTGTCTCTT GGTGTGCTC CTTTGTGAGG CCCAGATAG 1020  
10 ACAGGGAACA CTTGTGAGCA CACAGAGCAG CATCTGATGC CTTGTGTGT TTTGCAATGTG 1080  
CCCCCTGTCT ACTGACCAAT CAGTGTGTGCA TGAAGGCCAC GCCACCCAAA CTTTTCAGTT 1140  
TCCAAAGAGC TAGCCCTGCT CCAKCCAGTA CCATGTCTTA GCTGTGTGCT ATTGTATAGT 1200  
GTAATATATC TTTATGTATA ATAAATTTTT ATACCCAGC CATTCATGTA CTTTTCCTTG 1260  
TACTCTCCT TGTGGTTCCT TTGTCTGCT TGGCTGAAC CCAAAATGCT TTGGGTTGG 1320  
ACAGAGCTGG CTGAACCTTA GTTCTTCTAT CTATGAAATG GGAATATGAA TTAATCAGC 1380  
AGCTTTTAGG GCAGATTTGC CATGCAATAT ACAAGTAAAC TACCATAGTG CTCCTGGGT 1440  
25 ATTGCCATA TCCATATATT TCTGTGTAAA ATGAAGATAC TGATTTTATT GAG 1493

(2) INFORMATION FOR SEQ ID NO: 306:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 577 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

40 AATTGGCAG AGCAATATATA TACACTATAC TGGCAATTAC TGTTTACCC AGCCCGAAA 60  
GTGAGAGTG TATATTGGA AAATTACAC TCACTTACA TTGGTTCCA GAACTCTC 120  
ATAGCATT ACCCAGGAT CTACAGGAT GATGAGCA CTTATATG TTAATGACT 180  
CACTATATCT TATATTTTA TTGTTATTT TGTGTTTAT GCACAGTAC TTCAACCTT 240  
AAACTGCTT TGTATTGCT ATGTAACTT TTAAACATG CAGATCAGT TAGAAGTGT 300  
CATAGAGGA GAGCTAGAA TCCAGTACA TGAATTTTAA ATACCTGTC TTGTTTGG 360  
ATGTTAACA GTAAATGCA GTAGTAGCA AGAACAGT GATTATAT ACTATAGT 420  
AGGATTTCA TTTTAAATC ATCTTATCA AGATTAGAA CTCATTCCT GTGTTAAG 480  
55 GGAATGTTA ATTGCAAT AATCATTTCT GAACTAATG TTAATAAAA AAAAAAAAAA 540  
AAAAAAAAA AAAAAAAAAA AACTCGA 577

## (2) INFORMATION FOR SEQ ID NO: 307:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2860 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

5 GTGTGACCG CTCTCNCAT ATGGCTCCCC CGGCTGGCA GRWRTCTGT CNCRGTGCG 60  
TAGCCTTCC TCACAGGGA GAGTTAAGCT CCGTTCCTCC ACCGTCCCGG CTGGCAGGT 120  
GGGTGAGGG TGACCGAGAG ACCAGAACCT GCTTGTCTGA GCTTAGTGT CAGAGCTGG 180  
GAGGAGGTT CCGCGCTCC TCTGCTGCA GCGCGGCGAG CCGCTCCCGG CTTCACTTCC 240  
TCCCGAGCC CCGTCTACTG AGAAGCTCCG GCAATCCAGC AGCCGCCAG CCGTGGCTTC 300  
AGCTCCCGG GCTTCCATGC AGGCCAGAC CGAACGCCAC TGGGAGGAA GACAGGACC 360  
TTGACATTC CATCTGACA GAGTCTCTGG CTGGAACTGA CAGCTCTCT CTTCTAGGA 420  
TGACCTACC CTTCACTCT CCAGTTTTC GATTGAGAC ATTAGATGA GGCAGAGAG 480  
ATGGCTCTGA GCGGACAGA GGAAGCTGG ATTTTGGAG GGGCTCTCT CCGTGGAGT 540  
60 CACAGTTCCA GGGCAGGAC CGGAATTCG CCGCTTCAGA TAAGAGTCAA CTTCCACTA 660  
CCGAAGGGA ACAGTGTCCA GTACGCGGA TCACACCGA TTGACCGAG ATCGCTCTT 720  
CAATGGCTC TCCCGGGTG TCCCGAGGA TCTGGCTGGA CTTCCAGGT ACCTGAGCA 780  
GACGAGCAG TACCTCAGG ACTTGGAAA TACACAGAG GCTCCACAG TAAGACGGC 840  
TGATGAAGC TGTCTGAAA CTTTAAGAC GGGTCAATG CTTGCAATCT GCACTCTGT 900  
CAGATCGAC GGGACTCTGG CAATCTCTAG CCGCTGTAA ATGCCAGTG CACATGAC 960  
TATTAACCG GCGACAGGC TCTCACATC GCATTCAGA AAGAGAGTC TCGAGTGTG 1020  
GAGCTCTG GTGAGATG GGGCAATGT GCATGCCCGG GTCTGGCGG ACTTCTTCA 1080  
GAGGGGCAA GGGACTTGT TTTATTTGG TGAAGTACC CTCTCTTGG CCGTGTGAC 1140  
CAGAGTGG GATGTGTAA GCTACTCTT GGAAGACCA CACGAGCGG CCGAGCTCA 1200  
GGCAGTAC TCCAGGGA ACACAGTCT GCATGCCCTA GTGATGATC TCGGACAACT 1260  
CAGCTGAGA CATTGCACTG GTGACAGCA TGTATGATG GCTCTTCAA CTKGSGSC 1320  
SCCTTCTCC CTACCTTGA GCTTGAAGAC ATCCGACCC TGTAGAGT CAGGCTCTG 1380  
AAGCTGCC CCAAGGAGG CAAGTGTAG ATTTTCAGC ACATCTGCA CGGAGGTTT 1440  
TGAGGACTGA GCACTTTC CCGAAGTTC ACCGAGTGT GCTATGGGC TGTCCGGCTG 1500  
TGGCTGTAT ACTTGGCTTC TGTGACAGC TGTGAGAGA ACTCAGTGT GAGATCATT

1560 GCGTTCATTT GGAAGAGCCG GAGAGAGAGC GAAATGATGCG TTGTTGAGAGCG CGTGAAGAAA  
 1620 CTGTGTGAGG GGAATAGGGA TTGTGTGATC CCGAAGTTCT TGTAAACTT CCGTGTAAAT  
 1680 CTGATCTACA TGTTCATCTT CAGCGTGTCT GCGTCAATGC AGCTTAAGCT GAAAGAGAGC  
 1740 GCGGCGCGCTC AGCTGAAGAG GAGAGTTGGA AGCTGAATGC TGTGAAGAGG CCAAGATCTT  
 1800 AATCGTCAAG GGGGGAATCA CCGTCTGATG GGGCGAGCTG TGGTAATCTT GCGGCGCGCA  
 1860 CCGTGTCAATC TGAATCTGT TCAATGAGAG CTAACTTTGGA AATCTCTTTC CTGTTCAGAG  
 1920 CCGTCTCTCA CAGTGTGTGC CAGAGTGTGC TGTTCCTGTG GCGATGAGT GGTAACTGTGC  
 1980 CCGTCTGTGTG TTGTGCGGTG TGGATGAGCT GCGTGAAGCT GCTTAACTTA TACAGAGTGC  
 2040 GTTTCAGAGC AGAGAGAGTC TACAGTTTCA TGTATCTGGA AGCGTGTGTC AGCTGAAGCC  
 2100 AGAGAGCTTG GCGCGCGGAA GCTGTCTAGG GCGCGAATGC GAGAGATCA GTTGAAGCGA  
 2160 TGGAGAGGAA GAGAGAGAGG GCGAAGAGGG CCGATGAGAG GGTATATCTG GAAAGCTCTT  
 2220 TGGAGCTTT CAAATTCAGC ATTCGAGATG GCGAGCTGGC GTTTCAGAGC CAGCTGAGCT  
 2280 TTGCGGCGAT GGTGTGTGTG CTGTGTCTAG CCAAGAGCTT GGTGAAGTAC ATTCGTGTGC  
 2340 TGAAGATGCT CATTGCGGTC ATGAGAGGAA GGTGAAGATG TTGCGAAGTA CAGCTGAGAGC  
 2400 ATTCGAGAGC TGAAGAGAGC CATTGTGTGC CTGAGAGATG AGAATGAGTA TTGTGTGTGC  
 2460 AGGAGAGAGC AGCGGAGAGC TTGTATGTGT AGCGTTGTGA CTAAAGCGAG ATTCGAGAGC  
 2520 CCAATGAGAGC TTGTGTCTTCA GGTATGAGAA GGTGAAGATG GGTTCATATG GAGAGAGCCG  
 2580 TCGTAAGAGC GTTGTGAGAGC CCGTCAAGGG CAGTGTCTCC TCGAAGCTTC GAGAGAGCTG  
 2640 TCGTAAGCTTC CCGTCCAGAG GAGAGTGAAG ATGTGTGCTC TGAAGAGAGC TATGTGTGCG  
 2700 TCGAAGCTCT CCAATTCAGC TGAATGAGCA GATTCAGAGC GAGCGGAGAG GAGAGAGAGC  
 2760 AGAGATCTTTC CAGAGAGATC TCGTAAGCTT GGGTGTCCAG TGAATGTGCG TGGCAAAATAT  
 2820 AATATTTTAC TAACTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAATGAGG GGGGCGCGAT  
 2860 AAGCAATATC GCGCTATAG TGAATGCCAA TTAAGATAAA

(2) INFORMATION FOR SEQ ID NO: 308:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 876 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

60 CTGTGTGTGT CTGCGCTGTGT GCTGGCTGTG CTGAAGCTTC TTAACTATAC AGCTGTCTTC  
 120 CAGGAGAGAG GATATCTAGAG TTTCATGATC CAAAGAGCTT GATAGAGCTG AGCGAGAGAT  
 180 TTGCGCGCCC GAAAGCTCTTA CAGCGCGGAA TGGCAGAGAG TGAATGAGAG CCAATGAGAG  
 240 AGAGAGAGAG GAGGAGAGAG GGGCGGAGTA CAGGCGATATC CTGGAAGCTT CCGTTGAGAGT  
 300 CTTCAAATATC AGCATGTGGA TTGGGAGAGT GCGCTTCAGC GAGAGAGTGC AGTTCCGCGG  
 360 CATGTGTGTG CTGTGTGTGC TGGCTTAAGT GCTGTGTGAG TACATCTGTG TGTCTCAAGT  
 420 GGTATATGAG CTCAATAGAG GAGAGCGACA AGAGATGTGC CAGTGAAGAG TGAAGATCTT  
 480 GGAAGCTGCA GAAAGAGATC TTGTGTGTGAG AGATGAGAAA TGGCTATATG TGGTGAAGAA  
 540 AGAGAGAGAG GCGAGATGTG ATGCTGAAGC TTGTCAGTAA GCGAGATGCG AGCGCGAGAT  
 600 AGCGATGTGTC GTTCAAGGTG GAGAGAGTGA AGCTGAGCTTC ATTCGAGAGAG AGCGTCTCTTA  
 660 CCGTGTGTGA GGAAGCGTCA GGGCGAGTGT TTGCTTCAGC TTTCAGAGAG CCGTGTCTGTG  
 720 GTTCCGCTGC CAGAGAGATC GAGAGATGTG CCGTGTGAGG AATCTATATG CCGTGTCAAGC  
 780 TGTCTCAATC CAGCTGATGTC CCGAGATGCA GCGAGAGAGC AGAGAGAGAA GAGAGAGATC  
 840 TTTCAGAGCA CATTGTGTGT CTGTGTGTGT CCAATGAAAT CTGTGTGAGAA AATATATATT  
 876 TCACTAAAAA AAAAAAAAAA ACTGGA

(2) INFORMATION FOR SEQ ID NO: 309:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2025 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

60 CATAGAGCCG CTGATGCAAT CCGGAGAGC CTGTGATTC AGCGTAAGCT CTGTGATGAG  
 120 CAGCGAGAGC GCGTCCAGCA CAGAGAGAGG CAGCGAGAGC GCGTCCAGCA CAGAGAGAGG  
 180 CAGCGAGAGC GCGTTCAGCA CAGAGAGAGG GCGCGAGCTG GAGTGAAGCC CCAAGTGTGG  
 240 TGTGTGTGAG AAGAGAGCGG GCGCGAGTTC ATGAGAGTCT CCGTGTAGAG GCGTTCAGCA  
 300 GCTGCGAGCC GCGAGAGTGT ATTCCTTAGG TGGCGCGCTC CTGCGCGAGC CCGTTTGTGC  
 360 CCGTGCCTGC CATATGTGCG CTAACTGTGT AATTAATCAA AGCTTAATTT GTAAAGATGA  
 420 GCTGTGTGTG AGAGAAATGA GATCTATAGC GTGGTGTGCC TTGTCAAAGG GCGGATGTGCG  
 480 GTGAGAGAAA GAGAGAGAGC AAGCATGTGC GATGTGCATC CTCTTCATTT AAGCATGTGC  
 540 GGTGTGAGAG TTGCTGTGCC TTGCTGTAGAG AAGAGAAACT GCGAAAAACA AGAGCGGTAC

AGCAGACACT TCACAAAGCC AAGCCTAGGC GGCCTGTAGC ATCTGTGTTT AAGCGGTGC 600  
CTGTCTAGAA GCGCAGCGC CGACTTCGG TTCTCTCTTT AACTGAGAG AAGCTGATCC 660  
AGTTTCGGG AACAAATCC TTTTCTCAT TTGGGAGGG GGTATATAGT ACHTCAGGC 720  
ACCTCTTTA AACAGCCAA ACAGAGAGG GGAAGAGTG GGAATCATCT CGAGCTAGA 780  
GGCATTTGG ACACAAATC FAGGTAGTGA ACTTGAGAA ACCGATTTT AAGTTGGTG 840  
CATCTAGAA GCTTTGATG CAGACCAA CAAGCTTGA TTTTCTAGCA TCCTCTTAT 900  
GTCCAGCAA AGCAGGCRAC AAAATCTCTT GCTTTACAG ACAAATAT TTACGCAAC 960  
GTTCGGCATC ATGTGTTTG AAGGCTTAG TTCTGCTTTC TCCTCTCTCT CCACAGCCC 1020  
AAGCTCCAC CCTCATAGA TCAGCAGTG ATATATCTTG TTGAGGGAGA AGATCATTTA 1080  
GATTTGTTT GCATCTCTTA GATGAGGG CAACATCCA CAGCTGCTT GCTGTGATG 1140  
AGTGTCTTG CAGGGGGGG AGTAGAGCA CTGGGGTGG GCGGGAATG GGGTTACTG 1200  
ATGTAGGGA TTCTTGTG TGTTGTGAG ATCCAGTGA GTTGTGAT TTGTGATCC 1260  
CAGCTTGGT CAGGAAAT TTGTGTGATG GCTTAATCC AGTTTTCAT CTTCGACAG 1320  
TGCGCTGAA CGTGAACTCA GTAGCTGAC CTGTCTGAC CGGTCAAGTT CTGTGATCT 1380  
CAGAACTCT TTCTCTTGT GGGTGGGG TGGAACTCA GTGCGGAGC GGTGCTGAG 1440  
AAATGTAG GATTCGAAA TACATATCC ATGGGACTTT CTTTGCTCT CTGCTTCT 1500  
CTTTTCTGC TCCCTACT TTGCGGAT GGGGAGCAC CACTGAGTT TCTGGGGGC 1560  
CAGTGGGCT GCGAGTTCC TGTACTAGT CTTGTACTT TTGATTTTG CTCACCTGG 1620  
ATTTTCTCAT AGGAAGTTG GTCAGATGA ATTGATAT GTATGTCAG CACTGGGAC 1680  
CGAGGATTC TGGAGCCCG CAGTTGGAG GAGGAAGTAG TCGAGCTTC CAGGTGGGT 1740  
GAGAGGAT GACTGTTAC CTCGCGCA TCACCTTGA GCGCTTCTCT GCGCTTGGT 1800  
AGAAAAGTG GGAATCGGG CAGAGAGGC TGAGTAGGA TGGGAAC TAAGTGCACA 1860  
GTCTTTCCAG AGGAGTTCT TAATGAGTA TTTGTATTTA TTTCCAGC AATAAATTG 1920  
TACTTTTGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1980  
GAGGGGGGC GTTACCANT TCGCGTATA TGATCTTAA CACTC 2025

(2) INFORMATION FOR SEQ ID NO: 310:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3026 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi.) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

5 TAGGCAGAC TGAATATCC TAACCCCTA AGCTCCAGGT GCGCTGTGCG ACAGCAACT 60  
GGACTATAG AGGCTGGGC TGTGTCTTC TGTCTATAG CTCACCTTTT CCCCANAATC 120  
TTCTCTGGA GCTTTCAGC CAGGTGCTA AAGGATAG GTAGGAGCC TCTTCTATCT 180  
AATCTTAAA AGCAATATGT TGAACATCA TTCAACAGCT GATGCCCTAT AACCCCTGCC 240  
TGGATTTCTT CTTATTAGC TATPAAGAT AGCAAGATCT TTACATATTT CAGAGTGGTT 300  
TCAATGCTTT CTTACCTCTT CTATGCGCC CTCATTTTAT TTGACTAAG CATCACACAG 360  
TGGCACTAGC ATTATACCA GATATAGGA AATACAGTGC TTTATGCTTC TAACATTAAT 420  
GCTTTCAGTA TCAGGCTGC CTGAGAAAG GATGCGAGCC TCAGGGCTTC CTTATGTCTT 480  
CAACCAAG AGCTCCTTGA TGAAGTCAAT CTTTTCCTCC TATCTGTGTC TTCCCTCTCC 540  
GCTCTTAT GGTAGTGGG TACCCAGGCT GGTTCCTGGG CTAGGTAGTG GGGACCAAGT 600  
TCAATTAATC CTTATCAATT CTAGCAATGT AACATCAAGT ACCAGTGTTA GTGGAGAGAG 660  
CTGGGTTTTC CTAGTATACC CACTGCATCC TACTCTTACC TGTCTCAACC GCTGCTTCCA 720  
GGTATGAGC CTCTATAGTG TGGATTAAC TGATAGGGA GAGGAAATA CAGAGAGGC 780  
CTCTGTGTT CCGGCTCA GCGAGCTGC CACAGCCAT AACCAATTA AACAGAAATA 840  
CTGATCAT TTTTATCTG GGTTCCTTC ATTCCTACTG CACTTGTGTC TGTCTTGGCT 900  
GACTGGGAC ACCGCTTAC TACAGATCT CACAGAGAGA CTGGAGACTG TCCACTCTTA 960  
GCTCGAAT TACTGTGTA ATAACTTTC AGAACTGCTA CCATGAGTG AAAATGCCAC 1020  
ATTTTCTTT ATATTTCTA CCGATGTGG GAAAACTGG CTTTTCCTCA GCGCTTCCA 1080  
GGGCTTAAA CTCAGCCCT TCGATAGCA GTCCATCAG CTTATTTT TTTTAAAGAA 1140  
AAGTTCAGT TGTTTTCTT TTTACAGTTA CTCTCTCTT GCGCAAAAT TATTAAGCTT 1200  
AAGTGTAAA AAGGTCTTA ACACAGCTT CTGTCTTGT AAAATATGTA TTTATCAATCT 1260  
GTATTTTAA ATTCTGCTC TGAATAATGA CTGTCCATTT CTCACCTCAC TGCATTTGGG 1320  
GCGTTTCCA TTGCTCTGA TGTCTTTTAT CATTCAGGC CAGTGCAGAG AGGAGAGAGG 1380  
GAGACAGGG GTGCCAACA CTTGTGTTG TTTCTGACTG ATCTGAAACA AGAAGAGATA 1440  
ACACTGAGGC GCTGCTGCC ATGCAGACT CTCGAAACA CTTATCTGCC TCGAAGAGTG 1500  
GGCTTTCAG GGTCTTACT GGGAGCAGT TAAGCCCTT CCGCACTCT TCTTTTTTC 1560  
TTCTTACT CTTTGGCTT CAAAGATTT TGGAAAGAA ACAATATGCT TTACACTCAT 1620  
TTTCAATTC TAAATTTGA GGGGATCTG AAAATPACGG CAGGTGGCTT AAGGCTCTG 1680

	TTAAATGTAG GAGAGAGAA ACTTAAAGT TGAAGATTA AAAAGAGAT CCGTAACA	1740
	AAAAGAGT AGAAGTGTG TTGATTTTG CGACTTTCC TGTGTATAC AGCTACTAC	1800
5	CTGAGACAG TACATTTCA TTAAACAAAG AAGTGTGTG ACTGATCTG TGAAGATCG	1860
	AGTACTACAG CGACTCAAT CAGCTTACA GATTCACAG AGTCCACAG AATGCCACT	1920
	CGTTAAATG AGCTAGAAA ATTAACCTG GCAATGAG CAAGAGAAAT GAGAGAAAT	1980
10	CGATCTGTA GGTATAGGC AAGATGAAA GACAAAGAG GAAAAAGTA TCAAGAGAG	2040
	AAAAGAGAT ATTATTTTG GTCTGAAAG AAAAGCTTT GCTATCCAG ATGTACTCT	2100
	AGTACCTTA AGCATTTTG GTCCAGAAAT GCAAAAAA ATCAGCTATT GGTAAATTA	2160
15	TAATCTCCT TCCCTGAGT CAGTTTTTT AAAAAGTTA CTCTAGTTT TTACTTTTT	2220
	AATCTAAAA GAGAGGAGG CTGAGCCAT TCCCTTAGG AGTAAAGTA AAAAGATAG	2280
20	AAAAAGTTCA AAGCTTAAT AGATGACAG CTTCACAG TATTAAGCT AAAATTAGA	2340
	AGTCAATTA GCAAGCTGAG AAAATGATCT AGTCTGAT AGCTAACCA AGAGCAAGT	2400
25	ATTATTAAT TTGAATCCA AACTACTTTC TTAAATACAC TTGTGTCTCC ATTTTTCCA	2460
	GAGACAGAA TATGTCCCC CCAACTTTC TTGCTTCAA AATTAATC CAGATCCCA	2520
30	AGATCACTT ACAAATTAAT TTGCAAGAC ATCTCTTAC CCGATGCTT GTCTGAGCT	2580
	CAGCCAGCT CAGCCAAACA ACTGTGTCT GAAACCACT GCTTAAGCT TGTGGGAGG	2640
	GCGATTTAC TGAATAGGA GACCCAGAG TGAATGAAA AGGTGAGAG CTTCACAAAT	2700
35	TTGGCTCTC AGACTTGAT TGAAGCCA GACATGACA GCAAGAGAG ACTTGCCCA	2760
	GAAAAACCT GTGGTTGTG CTATTTCTG TCGAAGAAAT AGGTGAGCA GAGCTGTG	2820
40	GGTGCAATG AGAATTTGG AGCTGTGAT GTTGTATTC TCGACTGTG AATTTGTG	2880
	ATGTAAACA GATTTCTG TAACTTAAT GTCTGTAAA ATTAAGAGG TTAAACAGT	2940
	AAAAATTTCA ATTAAGATC AAAAAAAA AAAAAAAT CAGAGGGAG CCGGTATCC	3000
45	AATTTGCA ATTAAGATG TATTAC	3026

50 (2) INFORMATION FOR SEQ ID NO: 311:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 712 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

60 GAGGCTTGG TGTGACTTA CAGCTGGGT GAGCAGGTG CAGCAGCTT GTTCTCTTT

	CTCTCTGAG AACAGGCGT TTGTCTCCC GAGTTGAGAC TGTGAATAG TGTGGTCTT	120
	GTGTGTCTT CCAATCTAG CTCTCTCTG GTTGGAGCT TGTGAGCCA GACTAGAAA	180
5	CTGTGCTCT TGTGAAGTC GATGTGAGC TTGCACTTG GGGGCTTAC CTGTGAGCT	240
	GCTTGTGCT TCACTGTGA CAGCTGTGG GCGAGATAG AGCTGTGAC AATTTGAGA	300
10	GCTTGTGCT TGTGAAGCT ATGTGTGAG CAGCTTTGG GAGCTGTGT GACAGAGTC	360
	AGCTTACTG GGTATATGAG CTGAGCAG CTGGCCCCA GGGCTGTAG GCGAGAGCT	420
	AGACTCTCT GAGCAGCTG GAGCTGTGAG GAAAGCTGT GTTGGAGCT CTGAGAGGC	480
15	CTGTGTATG GGTTGGAGC AATCTCTGCT TTCTGTCTC TGTGTATCT CTGTGCTTT	540
	CGCTTCTGT ACTGTAGCT AACAGCCAG ACTTCTCT GAGCTGTAGT GCTGAGATG	600
20	TCAATTAAG CAGATGTCC TGTGGCCCA AAAAAAAA AAAAAAAA AAAAAAAA	660
	AAGTGAAGG GGGGCGGGT ACCGAATG CCGATATGA TGTAAACA TC	712

25 (2) INFORMATION FOR SEQ ID NO: 312:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1289 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

	CAAAATTTCA GAATTTGAG GAGCCAGAA AATTTATAG TGAAGATTT TGTCTTCCC	60
	TTGCCAAGT TCTGTATGAG CTGAGAAA AATTTATAG TGAAGATTT TGTCTTCCC	120
40	GTATTTGCT TTTTGTATG TTTTGTATG TTTTGTATG TTTTGTATG TTTTGTATG	180
	TCAACTGAA CTACATTTG AGAGAAAT TGTCTACATA GATATATTA TGAAGTGT	240
45	ACATATCTT GATGAGAAA AAAAATCTT GCAATCTTT AAGCAATAT GTTGTTTTC	300
	TGTGTGTTT TCCGTAGAG AAAATATAG TATTAAAGAG AACAAATAT ATTCAAGCT	360
50	TGAAGTAT TAAATATTC TTGCTGTGA TTATACATA TGTATATTT GCAAAATAT	420
	GCTTTTTTA AGGAAAGTA TAAATGATA CATATGAAA AAGGATAG TGAACCTTT	480
	GTCTCTCTC CAGTGAAGT AACAAGAGC ATGTATATC ATGTCTTAC ACTTCTTCT	540
55	TCTTATCTG TTAATAGCT TTTATTAAT TTGAGAGAG ACAGAGATG GCGAAGGGA	600
	AGAGAGACT TATTGACTA ACAGAGCTT GTTGTATCA CAGAGCTTT GAGCTGTGAG	660
60	GAGGCTCTC AATGACAGG GCGCCAGAG GAGAGAGAA GTGGGGAAA GCTTGGCTCTC	720



5  
10  
15  
20

GGCGCTGGGG AGCTTTTGCCA TGTGAGCCAC GCGCTCTCCA GCGCATGCTC CTTCAGACTTG 780  
GAAATGTCAA CCGAGGCCCT TACACGAGCC CTCAGGATC TATATAGACTT GATCTACTC 840  
TAAAGGATA TTTATATCAA CCTCACTACA TTGTAGTCA GTCCAAGCAG TAACCTGAA 900  
ATGGGGGTGT TCCGACTTTC AGCAGATGG CCAGCGGTC CCGTGGGGC TGTGGAGCG 960  
GGCTTATCCT TGTCTGTGCG CAACCTGCG GTCCGACTTC CTCGCGCCCG ATGGGGTAC 1020  
CCGCTCCCTG TGTGTGTG TGCTAGTGTG TGAGTCCAGC TAAAGAGCA AACAGAGCC 1080  
CGTGGGCCCA GCTGGGAAGG TCGCTGGACA AGGCTCCGAC GTCTCCGAG TCGAGCCCTT 1140  
GGATGGGCAAT TCGTTTGTGT GCGTTATGCC TGGAGAAATCT GTATACGGCT CCGCTATAGA 1200  
AATATAGCCT CTTCATGCTG TATTAAAGG ACTTTTAAA GCAAAAAA AAAAANA 1260  
CTTGAGGGGG GAGCCGTAC CCAATTTC 1289

## (2) INFORMATION FOR SEQ ID NO: 313:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Ile Leu Phe Ser Ser Ser 15  
1 5 10

35  
Leu Pro Phe Leu Trp Leu 20

## (2) INFORMATION FOR SEQ ID NO: 314:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 128 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

50  
Met Met Phe Leu Thr Gln Gly Gly Pro Leu Pro Ser Thr Arg Ala Arg 15  
1 5 10  
Pro Thr Cys Gln Ala Gly Ala Leu Pro Lys Pro Ser Gly Leu Leu Gly 30  
20 25

55  
Val Thr Cys Trp Asn Gly Leu Lys Gly Pro Leu Cys Gly Asn Arg Cys 45  
35 40

Ser Pro Asn Thr Leu Leu Ala Ala Arg Gln Ala Leu Trp Lys Gly 50  
50 55

60  
Arg Gly Arg Thr His Gln Asp Leu Pro Gly Pro Leu Gln Gly Arg Gln

65  
70  
75  
80

Leu Gly Pro Gln Pro Lys His Leu Ala Leu Leu Pro Pro Arg Gly Gln 95  
85  
5  
Glu Ala Ser Trp Ala Ser Ser Leu Pro Gly Gln Gly Pro Leu Pro Leu 110  
100 105  
10  
Pro His Ile Asn Cys Thr Val Phe Ser Leu Lys Ala Ser Phe Ile Lys 125  
115 120

## (2) INFORMATION FOR SEQ ID NO: 315:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

25  
Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu 15  
1 5 10

30  
Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser 25  
20

## (2) INFORMATION FOR SEQ ID NO: 316:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

40  
Met Asp Gly Phe Ser Ser Arg Leu Phe Ser Ser Leu Pro Phe Val Ala 15  
1 5 10

45  
Leu Gln Trp Phe Ile Val Ile Ser His Leu Leu Ser Leu Ser Leu Ser 30  
20 25

Ala Cys Cys Tyr Gln Thr His Cys Ser Leu Xaa Gln Leu Ser Ser Ala 45  
35 40

50  
Phe Ser Xaa Met Gly Glu Ser Cys Val Gly Glu Arg Glu Tyr Xaa Phe 60  
50 55

## (2) INFORMATION FOR SEQ ID NO: 317:

(1) SEQUENCE CHARACTERISTICS:

547

(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

5 Met Pro Leu Ile Asn Leu Leu Leu Leu Tyr Tyr Val Pro Asn Gly Gly  
1 5 10 15

10 Lys Gln Asp Lys Lys  
20

(2) INFORMATION FOR SEQ ID NO: 318:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

Met Gly Arg His Leu Val Leu Val Met Phe Ile Thr Thr Ser Leu His  
1 5 10 15

25 Ser Gly Thr Pro Val Pro Glu Asn Val Ile Cys Gly Val Thr Lys Gly  
20 25 30

Pro Gln Gly Lys Lys Lys  
35

(2) INFORMATION FOR SEQ ID NO: 319:

35 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

40 Met Leu Trp Trp Ser Arg Asp Tyr Thr Met Val Phe Leu Leu Phe Thr  
1 5 10 15

45 Met Val Phe Thr Gly Asp Leu Val Ile Arg Gly Arg Thr Glu Leu Ser  
20 25 30

Leu

50 (2) INFORMATION FOR SEQ ID NO: 320:

55 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 88 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

60 Met Val Cys Ser Ser Leu Cys Asp Ile Gly Gly Ile Ile Thr Pro Phe

548

1 5 10 15

Ile Val Phe Arg Leu Arg Glu Val Trp Gln Ala Leu Pro Leu Ile Leu  
20 25 30

5 Phe Ala Val Leu Gly Leu Leu Ala Ala Gly Val Thr Leu Leu Leu Pro  
35 40 45

10 Glu Thr Lys Gly Val Ala Leu Pro Glu Thr Met Lys Asp Ala Glu Asn  
50 55 60

Leu Gly Arg Lys Ala Lys Pro Lys Glu Asn Thr Ile Tyr Leu Lys Val  
65 70 75 80

15 Gln Thr Ser Glu Pro Ser Gly Thr  
85

20 (2) INFORMATION FOR SEQ ID NO: 321:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

25 Met Gln Pro Gly Ala Gly Val Leu Val Leu Gly Leu Leu Leu Pro Pro  
1 5 10 15

30 Pro Gln Ser Pro Ser Leu Ser  
20

35 (2) INFORMATION FOR SEQ ID NO: 322:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

45 Met Thr Phe Thr Leu Gly Asp Ser Gln Val Leu Leu Ile Asn Leu Phe  
1 5 10 15

Pro Ser Met Pro Ser Gly Ser Cys Ala Arg Pro  
20 25

50 (2) INFORMATION FOR SEQ ID NO: 323:

55 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

60 Met Cys Leu Glu Cys Trp Ala Glu Asn Leu Leu Pro His His Thr Ser

549

1 5 10 15  
Ser Leu Leu Asn Pro Arg His Leu Pro Ser Ile Pro Ala Met Phe Pro 20 25 30  
5 Val Ser Ser Gly Cys Phe Gln Glu Gln Gln Glu Met Asn Lys Ser Leu 35 40 45  
10 Val Ser Cys Leu Phe Val Leu His Phe Val Leu His Cys Ile Phe Xaa 50 55 60  
15 (2) INFORMATION FOR SEQ ID NO: 324:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 196 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:  
20 Met Leu Ser Thr Ser Glu Tyr Ser Gln Ser Pro Lys Met Glu Ser Leu 1 5 10 15  
Ser Ser His Arg Ile Asp Glu Asp Gly Glu Asn Thr Gln Ile Glu Asp 20 25 30  
30 Thr Glu Pro Met Ser Pro Val Leu Asn Ser Lys Phe Val Pro Ala Glu 35 40 45  
35 Asn Asp Ser Ile Leu Met Asn Pro Ala Gln Asp Gly Glu Val Gln Leu 50 55 60  
Ser Gln Asn Asp Asp Lys Thr Lys Gly Asp Asp Thr Asp Thr Arg Asp 65 70 75 80  
40 Asp Ile Ser Ile Leu Ala Thr Gly Cys Lys Gly Arg Glu Glu Thr Val 85 90 95  
Ala Glu Glu Val Cys Ile Asp Leu Thr Cys Asp Ser Gly Ser Gln Ala 100 105 110  
45 Val Pro Ser Pro Ala Thr Arg Ser Glu Ala Leu Ser Ser Val Leu Asp 115 120 125  
Gln Glu Glu Ala Met Glu Ile Lys Glu His Phe Pro Glu Glu Gly Ser 130 135 140  
50 Ser Gly Ser Glu Val Glu Glu Ile Pro Glu Thr Pro Cys Glu Ser Gln 145 150 155 160  
55 Gly Glu Glu Leu Lys Glu Glu Asn Met Glu Ser Val Pro Leu His Leu 165 170 175  
Ser Leu Thr Glu Thr Gln Ser Gln Gly Leu Cys Leu Arg Arg His Pro 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995

550

Lys Lys Lys Lys 195  
5 (2) INFORMATION FOR SEQ ID NO: 325:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 352 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:  
10 Met Gly Gly Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg Lys 1 5 10 15  
Arg Leu Leu Glu Gln Glu Lys Ser Leu Ala Gly Trp Ala Leu Val Leu 20 25 30  
20 Ala Xaa Xaa Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp 35 40 45  
Phe Gly Gly Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr 50 55 60  
25 Leu Trp Leu Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val 65 70 75 80  
30 Val Pro Gly Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val 85 90 95  
Met Gly Val Cys Cys Thr Ala Leu Leu Val Ala Val Ala Arg Lys 100 105 110  
35 Leu Glu Phe Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp 115 120 125  
Ile Gln Tyr Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln 130 135 140  
40 Glu Ala Trp Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala 145 150 155 160  
Ala Arg Xaa His Gln Arg Xaa Leu Leu Ala Ala Ile Asn Ala Phe Arg 165 170 175  
45 Gln Val Arg Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met 180 185 190  
50 Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Asn 195 200 205 210  
Leu Ser Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu Ala 215 220 225  
55 Gly Lys Leu Asp Ala Leu Thr Glu Leu Ser Thr Ala Leu Gly Pro 230 235 240  
Arg Gln Leu Pro Glu Pro Ser Gln Gln Ser Lys Xaa 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995

5 (2) INFORMATION FOR SEQ ID NO: 326:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val  
1 5 10 15

15 Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala  
20 25 30

Val Ile Lys Asn Asn Ser His Tyr Gln Thr Ser Lys Ala Leu Gln Leu  
35 40 45

20 Gln Lys Thr Thr Gln Asn Lys Gln Ser Asn Pro Phe Ile Leu Gln Val  
50 55 60

25 Asn Lys Leu Xaa  
65

30 (2) INFORMATION FOR SEQ ID NO: 327:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

Met Gly Gln Gly Lys Asn Gly Phe Gly Gly Phe Val His Thr Ala Asp  
1 5 10 15

40 Ala Cys Trp Gln Gly Val His Ser Gln Pro Val Cys Arg Thr Val His  
20 25 30

Thr Val His Thr Cys His His Gln Ala Phe Leu Val Leu Ile Gly Trp  
35 40 45

45 Ser Lys Ser Gly Lys Gln Arg Lys Gln Ala Phe Leu Thr Ala Ile Ile  
50 55 60

50 Leu Asn Ser Arg Ser Ile His Ile Ser Cys Ser Trp Pro Pro Ser Pro  
65 70 75 80

Val Pro Gln Xaa

55 (2) INFORMATION FOR SEQ ID NO: 328:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

5 Met Leu Leu Ile Asn Leu Leu Trp Leu Val Thr Met Ile Lys Ser Val  
1 5 10 15

Ile Asn Asn Asn Ile Ile Leu Phe Leu Lys Lys Lys Ser Leu Phe Phe  
20 25 30

10 Ile Asp Ser Val  
35

15 (2) INFORMATION FOR SEQ ID NO: 329:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 63 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

Met Thr Phe Pro Phe Gln Lys Lys Ile Val Ala Phe Ser Ala Phe Tyr  
1 5 10 15

Leu Ile Pro Gly Gln Ser Arg Leu Ala Pro Thr Phe Asn Pro Ser Ala  
20 25 30

30 Asp Met Thr Val Ile Leu Arg Gly Arg Ala Gln His Lys Thr Ala Met  
35 40 45

Leu Gln Ser Tyr Asn Trp Lys Val Ser Cys Gln Leu Arg Gln Xaa  
50 55 60

35 (2) INFORMATION FOR SEQ ID NO: 330:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

Met His Ser Lys Gly Ser Ser Leu Leu Leu Phe Leu Pro Gln Leu Ile  
1 5 10 15

50 Leu Ile Leu Pro Val Cys Ala His Leu His Gln Gln Leu Asn Cys  
20 25 30

Phe His Arg  
35

55 (2) INFORMATION FOR SEQ ID NO: 331:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

553

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:  
5 Met Gly Ala Leu Val Leu Leu Cys Leu Leu Val Gly Val Gln Gln 15  
1 5 10  
Ser Gly Ser Val Trp Asp Ser 20

10

(2) INFORMATION FOR SEQ ID NO: 332:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:  
20 Met Gln Ser Ala Glu Ile Leu Ser Trp Trp Asp Val Leu Leu His Asp Phe 15  
1 5 10  
Leu Phe Ser Leu Phe Leu Trp Pro Ala Phe Glu Asp Arg Ala Leu Leu 30  
25 20  
Ile Phe Thr Leu Asn Gln Ile Val 40  
35

30

(2) INFORMATION FOR SEQ ID NO: 333:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 111 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:  
40 Met Gln Ser Leu Val Gln Trp Gly Leu Asp Ser Tyr Asp Tyr Leu Gln 15  
1 5 10  
Asn Ala Pro Pro Gly Phe Phe Pro Arg Leu Gly Val Ile Gly Phe Ala 30  
20 25  
Gly Leu Ile Gly Leu Leu Ala Arg Gly Ser Lys Ile Lys Lys Leu 45  
35 40  
Val Tyr Pro Pro Gly Phe Met Gly Leu Ala Ala Ser Leu Tyr Tyr Pro 60  
50 55  
Gln Gln Ala Ile Val Phe Ala Gln Val Ser Gly Glu Arg Leu Tyr Asp 80  
65 70 75  
Trp Gly Leu Arg Gly Tyr Ile Val Ile Glu Asp Leu Trp Lys Glu Asn 95  
85 90  
Phe Gln Lys Pro Gly Asn Val Lys Asn Ser Pro Gly Thr Lys Xaa 110  
100 105

60

(2) INFORMATION FOR SEQ ID NO: 334:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 106 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:  
10 Met Ala Pro Ser Leu Leu Leu Ala Pro Leu Cys Ser Leu Glu Ala 15  
1 5 10  
Val Leu Ser Ser Pro Leu Glu Lys Gln Cys Gln Leu Pro Gly Ile Phe 30  
20 25  
Cys Gln Leu Gln Leu Pro Cys Pro Leu Leu Ser Ala Gln Leu Leu 45  
35 40  
Lys Gly Ile Val Xaa Pro Arg Cys Pro Ala Ser Leu Pro Gln Pro Pro 60  
50 55  
His Pro Ala Pro Ser Trp His Leu Pro Leu His Cys Thr Glu Arg Xaa 80  
65 70 75  
Pro His His Leu Pro Leu Gln Gly Ser Ser Asn Met Glu Glu Xaa 95  
85 90

30

(2) INFORMATION FOR SEQ ID NO: 335:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 50 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:  
40 Met Thr Thr Cys Leu Phe Gly Leu Leu Ser Cys Glu Met Ser Ala Gln 15  
1 5 10  
Val Ser Gln Lys Ser Cys Val Tyr Asp Glu Ser Glu Cys Phe Ser Ser 30  
20 25  
Val Gly Gln Leu Leu Ala Leu Leu Ile Leu Val Tyr Val Leu Pro Ser 45  
35 40  
Ile Xaa 50  
50

50

(2) INFORMATION FOR SEQ ID NO: 336:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
55  
60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

5 Met Leu Trp Lys Cys Ser Gln Asn Ile Ala Arg Cys Leu Leu Leu Leu  
1 5 10 15  
Leu Ala Leu Val Glu Ile Lys Leu Glu Asp Leu Leu Gln Ser Gln Leu His  
20 25 30  
10 Pro Thr Trp Lys Ser Ile Pro Gly Pro Ser Pro Arg Asn Gln His Arg  
35 40 45

15

(2) INFORMATION FOR SEQ ID NO: 337:

20 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

25 Met Leu Ile Pro Leu Gln Cys Leu Phe Ser Ser Asp Arg Met Leu Thr  
1 5 10 15  
Phe Leu Thr Pro Trp Gln Lys Gly Glu Lys Cys Val Leu Gly Trp Val  
20 25 30  
Thr Lys Phe Leu Ser Glu Ile Ser Xaa  
35 40

35

(2) INFORMATION FOR SEQ ID NO: 338:

40 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

45 Met Thr Phe Ser Ser Leu Lys Leu Phe Val Leu Thr Thr Cys Ile Ile Lys  
1 5 10 15  
Gly Leu Glu Arg Phe Ile Ile Leu Arg Glu Val Cys Asn Gln Glu Ile  
20 25 30  
50 Gln Arg Ser Leu Ser Ser Asn Leu Val His Val Leu Leu Gln Pro Ala  
35 40 45  
Thr Phe Lys Asp Val Leu Val Thr Glu Ile Ile Cys Leu Cys Met Cys  
55 60  
Leu Tyr Ser Ile Lys Tyr Met Pro Pro Gln Lys Lys  
65 70 75

60

(2) INFORMATION FOR SEQ ID NO: 339:

5 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

10 Lys Val Tyr Ile Phe Leu Ile Phe Met Val Leu Ile Leu Pro Ser Leu  
1 5 10 15  
Gly Leu Thr Arg Tyr Met Pro Pro Xaa Ser Xaa Leu Asn Ser Glu  
20 25 30

15

(2) INFORMATION FOR SEQ ID NO: 340:

20 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

25 Met Ala Lys Ile Ser Pro Phe Glu Val Val Lys Arg Thr Ser Val Pro  
1 5 10 15  
Val Leu Val Gly Leu Val Ile Val Ile Val Ala Thr Glu Leu Met Val  
20 25 30  
Pro Gly Thr Ala Ala Val Thr Gly Lys  
35 40

35

(2) INFORMATION FOR SEQ ID NO: 341:

40 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

45 Met Arg Leu Phe Phe Ile Gly Phe Leu Leu Leu Phe Ser Phe Gly Leu  
1 5 10 15  
Leu Arg Gln Pro Ser Leu Ser Ala Glu His  
20 25

50

(2) INFORMATION FOR SEQ ID NO: 342:

55 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

60

Met Val Phe Ser Val Ser Ala Leu Ala Leu Leu Met Leu Leu  
1 5 10 15

Arg Ser Ser Asp Leu Ala Lys Lys Thr Glu  
20 25

(2) INFORMATION FOR SEQ ID NO: 343:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 157 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

Met Ser Leu Glu Phe Tyr Gln Lys Lys Lys Ser Arg Trp Pro Phe Ser  
1 5 10 15

Asp Glu Cys Ile Pro Trp Glu Val Trp Thr Val Lys Val His Val Val  
20 25 30

Ala Leu Ala Thr Glu Gln Glu Arg Gln Ile Cys Arg Glu Lys Val Gly  
35 40 45

Glu Lys Leu Cys Glu Lys Ile Ile Asn Ile Val Glu Val Met Asn Arg  
50 55 60

His Glu Tyr Leu Pro Lys Met Pro Thr Gln Ser Glu Val Asp Asn Val  
65 70 75 80

Phe Asp Thr Gly Leu Arg Asp Val Gln Pro Tyr Leu Tyr Lys Ile Ser  
85 90 95

Phe Gln Ile Thr Asp Ala Leu Gly Thr Ser Val Thr Thr Thr Met Arg  
100 105 110

Arg Leu Ile Lys Asp Thr Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly  
115 120 125

Ser Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu  
130 135 140

Gly Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp Xaa  
145 150 155

(2) INFORMATION FOR SEQ ID NO: 344:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 520 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

Met Phe Leu Leu Pro Leu Pro Ala Ala Gly Arg Val Val Val Arg Arg  
1 5 10 15

Leu Ala Val Arg Arg Phe Gly Ser Arg Ser Leu Ser Thr Ala Asp Met

Thr Lys Gly Leu Val Leu Gly Ile Tyr Ser Lys Glu Lys Glu Asp Asp  
35 40 45

Val Pro Gln Phe Thr Ser Ala Gly Glu Asn Phe Asp Lys Leu Leu Ala  
50 55 60

Gly Lys Leu Arg Glu Thr Leu Asn Ile Ser Gly Pro Pro Leu Lys Ala  
65 70 75 80

Gly Lys Thr Arg Thr Phe Tyr Gly Leu His Gln Asp Phe Pro Ser Val  
85 90 95

Val Leu Val Gly Leu Gly Lys Lys Ala Ala Gly Ile Asp Glu Gln Glu  
100 105 110

Asn Trp His Glu Gly Lys Glu Asn Ile Arg Ala Ala Val Ala Gly  
115 120 125

Cys Arg Gln Ile Gln Asp Leu Glu Leu Ser Ser Val Glu Val Asp Pro  
130 135 140

Cys Gly Asp Ala Gln Ala Ala Glu Gly Ala Val Leu Gly Leu Tyr  
145 150 155 160

Glu Tyr Asp Asp Leu Lys Gln Lys Lys Lys Met Ala Val Ser Ala Lys  
165 170 175

Leu Tyr Gly Ser Gly Asp Gln Glu Ala Trp Gln Lys Gly Val Leu Phe  
180 185 190

Ala Ser Gly Gln Asn Leu Ala Arg Gln Leu Met Glu Thr Pro Ala Asn  
195 200 205

Glu Met Thr Pro Thr Arg Phe Ala Glu Ile Ile Glu Lys Asn Leu Lys  
210 215 220

Ser Ala Ser Ser Lys Thr Glu Val His Ile Arg Pro Lys Ser Trp Ile  
225 230 235 240

Glu Glu Gln Ala Met Gly Ser Phe Leu Ser Val Ala Lys Gly Ser Asp  
245 250 255

Glu Pro Pro Val Phe Leu Glu Ile His Tyr Lys Gly Ser Pro Asn Ala  
260 265 270

Asn Glu Pro Pro Leu Val Phe Val Gly Lys Gly Ile Thr Phe Asp Ser  
275 280 285

Gly Gly Ile Ser Ile Lys Ala Ser Ala Asn Met Asp Leu Met Arg Ala  
290 295 300

Asp Met Gly Gly Ala Ala Thr Ile Cys Ser Ala Ile Val Ser Ala Ala  
305 310 315 320

Lys Leu Asn Leu Pro Ile Asn Ile Ile Gly Leu Ala Pro Leu Cys Glu  
325 330 335

Asn Met Pro Ser Gly Lys Ala Asn Lys Pro Gly Asp Val Val Arg Ala

340 345 350  
Lys Asn Gly Lys Thr Ile Gln Val Asp Asn Thr Asp Ala Glu Gly Arg  
355 360 365  
5 Leu Ile Leu Ala Asp Ala Leu Cys Tyr Ala His Thr Phe Asn Pro Lys  
370 375 380  
10 Xaa Ile Leu Asn Ala Ala Thr Leu Thr Gly Ala Met Asp Val Ala Leu  
385 390 395 400  
Gly Ser Gly Ala Thr Gly Val Phe Thr Asn Ser Ser Tyr Leu Tyr Asn  
405 410 415  
15 Lys Leu Phe Glu Ala Ser Ile Glu Thr Gly Asp Arg Val Tyr Arg Met  
420 425 430  
Pro Leu Phe Glu His Tyr Thr Arg Gln Val Val Asp Cys Gln Leu Ala  
435 440 445  
20 Asp Val Asn Asn Ile Gly Lys Tyr Arg Ser Ala Gly Ala Cys Thr Ala  
450 455 460  
Ala Ala Phe Leu Lys Glu Phe Val Thr His Pro Lys Tyr Ala His Leu  
465 470 475 480  
25 Asp Ile Ala Gly Val Met Thr Asn Lys Asp Glu Val Pro Tyr Leu Arg  
485 490 495  
30 Lys Gly Met Thr Gly Arg Pro Thr Arg Thr Leu Ile Glu Phe Leu Leu  
500 505 510  
Arg Phe Ser Gln Asp Asn Ala Xaa  
515 520  
35 (2) INFORMATION FOR SEQ ID NO: 345:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 345:  
45 Thr Ile Leu Phe Leu Phe Leu Gln Leu Ser Ala Leu Arg Leu Ile Val  
1 5 10 15  
Gly Lys Asp Ser Ile Asp Ile Asp Ile Ser Ser Arg Arg Arg Glu Asp  
20 25 30  
Gln Ser Leu Arg Leu Asn Ala  
35

(2) INFORMATION FOR SEQ ID NO: 346:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 234-amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 346:  
5 Met Thr Ser Glu Leu Asp Ile Phe Val Gly Asn Thr Thr Leu Ile Asp  
1 5 10 15  
Glu Asp Val Tyr Arg Leu Tyr Leu Asp Gly Tyr Ser Val Thr Asp Ala  
20 25 30  
10 Val Ala Leu Arg Val Arg Ser Gly Ile Leu Glu Gln Thr Gly Ala Thr  
35 40 45  
15 Ala Ala Val Leu Gln Ser Asp Thr Met Asp His Tyr Arg Thr Phe His  
50 55 60  
Met Leu Glu Arg Leu Leu His Ala Pro Pro Lys Leu Leu His Gln Leu  
65 70 75 80  
20 Ile Phe Gln Ile Pro Pro Ser Arg Gln Ala Leu Leu Ile Glu Arg Tyr  
85 90 95  
Tyr Ala Phe Asp Glu Ala Phe Val Arg Glu Val Leu Gly Lys Lys Leu  
100 105 110  
25 Ser Lys Gly Thr Lys Lys Asp Leu Asp Asp Ile Ser Thr Lys Thr Gly  
115 120 125  
Ile Thr Leu Lys Ser Cys Arg Arg Gln Phe Asp Asn Phe Lys Arg Val  
130 135 140  
30 Phe Lys Val Val Glu Glu Met Arg Gly Ser Leu Val Asp Asn Ile Gln  
145 150 155 160  
35 Gln His Phe Leu Leu Ser Asp Arg Leu Ala Arg Asp Tyr Ala Ala Ile  
165 170 175  
Val Phe Phe Ala Asn Asn Arg Phe Glu Thr Gly Lys Lys Lys Leu Gln  
180 185 190  
Tyr Leu Ser Phe Gly Asp Phe Ala Phe Cys Ala Glu Leu Met Ile Gln  
195 200 205  
45 Asn Tyr Thr Leu Gly Pro Val Asp Ser Gln Met Asp Asp Met Met  
210 215 220  
Asp Leu Asp Arg Asn Phe Ser Arg Thr Xaa  
225 230 235  
50 (2) INFORMATION FOR SEQ ID NO: 347:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 169 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(K1) SEQUENCE DESCRIPTION: SEQ ID NO: 347:  
60 Met Ala Ala Ala Val Ala Gly Met Leu Arg Gly Gly Leu Leu Pro Gln



1 5 10 15  
 Ala Gly Arg Leu Pro Thr Leu Cln Thr Val Arg Tyr Gly Ser Lys Ala 30  
 20 25 30  
 5 Val Thr Arg His Arg Arg Val Met His Phe Cln Arg Gln Lys Leu Met 45  
 35 40 45  
 10 Ala Val Thr Glu Tyr Ile Pro Pro Lys Pro Ala Ile His Pro Ser Cys 60  
 50 55 60  
 Leu Pro Ser Pro Pro Ser Pro Gln Glu Glu Ile Gly Leu Ile Arg 75  
 65 70 80  
 15 Leu Leu Arg Arg Glu Ile Ala Val Phe Cln Asp Asn Arg Met Ile 95  
 85 90 95  
 Ala Val Cys Cln Asn Val Ala Leu Ser Ala Glu Asp Lys Leu Leu Ile 110  
 100 105 110  
 20 Ala Thr Pro Ala Ala Glu Thr Cln Asp Pro Asp Glu Gly Leu Pro Gln 125  
 115 120 125  
 25 Pro Gly Pro Glu Ser Pro Ser Trp Arg Ile Pro Ser Thr Lys Ile Cys 140  
 130 135 140  
 Cys Pro Phe Leu Trp Gly Thr Thr Cys Cys Trp Ser Val Lys Ser Pro 155  
 145 150 160  
 30 Arg Ser Arg Arg Trp Tyr Gly Ser Xaa 165  
 160 165

35 (2) INFORMATION FOR SEQ ID NO: 348:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

Met Lys Arg Ser Phe Leu Leu Pro Leu Leu Leu Val Gly Phe Leu Asp 15  
 1 5 10 15  
 45 Thr Ala His Leu Ile Leu Leu Glu Thr Leu Ser Val Cys Leu Trp Leu 30  
 20 25 30  
 Pro Ser Leu Ile Asp Ser Arg Cys Val Met Ser 40  
 35 40 45

(2) INFORMATION FOR SEQ ID NO: 349:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

Met Lys Glu Gly Pro Pro Cys Lys Arg His His Tyr Tyr Gln Asn Cys 15  
 1 5 10 15  
 5 Gly Ala Lys Leu Leu Val Ser Leu Phe Gly Glu Thr Asn Gln Ile His 30  
 20 25 30  
 Leu Leu Glu Thr Gln Val Gly Thr Glu Lys Gly Gly Arg Ile Trp 45  
 35 40 45  
 10 Glu Glu Lys Trp Arg Ile Ser Ser Thr Val Leu Phe Ile Ser Val Asn 60  
 50 55 60  
 Ser Tyr Val Glu Gly Ser Val Leu Glu Ile Lys Leu Phe Tyr 75  
 65 70 75  
 15 (2) INFORMATION FOR SEQ ID NO: 350:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

Met Ser Glu Ile Leu Ser Leu Leu Phe Cys Leu Leu Gly Pro Ala Leu 15  
 1 5 10 15  
 30 Asp Glu Arg Arg Glu Glu Lys Asp 20  
 20 25

35 (2) INFORMATION FOR SEQ ID NO: 351:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

Met Ser Ser Ala Gly Thr Ala Thr Pro Leu Glu Met Asp His Lys Leu 15  
 1 5 10 15  
 45 Thr Ser Gln Pro Gly Arg Pro Ser Phe Tyr Cys Asn Ser Arg His Ser 30  
 20 25 30  
 Ile Val Gly Ser Ser His Gln Leu Gly Phe Trp Phe Ser His Leu Glu 45  
 35 40 45  
 Ser Ser Gly Leu Lys Val Phe Gln Val Ser Leu Pro Cys Glu Cys Val 60  
 50 55 60  
 Asn Leu Pro Thr Arg Ile Ala Ser Val Val Leu Ser Leu Met Ser Leu 80  
 65 70 75  
 Leu Val Val Gly Gln Ala Pro Ala Trp Glu Gly Ser Leu Leu Arg Gly 95  
 85 90 95  
 60

Arg Pro Ala Gly Gly Ala His Leu Cys Ala Met Xaa Val Ile Glu Gly  
100 105 110

Leu Val Val Asp Val Gly Glu Arg Ile Leu His Gly Gln Arg Glu Val  
115 120 125

Gly Gln Val Ser Gln Val Leu Pro Ala Leu Ser Leu Gly Leu Val Phe  
130 135 140

Leu Cys Gln Gly Thr Val Glu Lys Val Ser Gly Ala Ala His Cys Ser  
145 150 155 160

Ser Leu Leu Cys Cys Leu Pro Trp Gln Cys Ser Gly Gly Gly Phe Pro  
165 170 175

Thr Xaa Arg Cys Ser Arg Pro Tyr Phe Ser Ser His Lys Gly Val Ala  
180 185 190

Ala Thr Leu Ala Leu Thr Cys His Cys Asp Lys Val His Val Ala Gly  
195 200 205

Leu Gly Lys Asp Trp Ala Ile Glu Gln Arg Arg Thr Cys Glu Ser  
210 215 220

Asp Xaa Glu Xaa Xaa Pro Phe Thr Leu Ala Gly Leu Val Leu Val Leu  
225 230 235 240

Arg Phe Cys Gln Val Val Leu Val Trp Ile Pro Gln Leu Gly Asp Lys  
245 250 255

His Trp Arg Gly Met Thr Arg Leu Gly Arg Val Ser Leu Thr Ser Ser  
260 265 270

Ile Xaa

(2) INFORMATION FOR SEQ ID NO: 352:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

Met Ile Phe Thr Ser Val Thr Lys Gly Ile Leu Leu Ile Ala Leu Trp  
1 5 10 15

Val Pro Leu Phe His Phe Met Leu Ile Asp Ser Ile Leu Gly Pro Ser  
20 25 30

Arg Leu Leu Thr Asp Gly Val Pro Phe Asn Pro Trp His Val Xaa  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 353:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

Met Lys Thr  
1

(2) INFORMATION FOR SEQ ID NO: 354:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

Met Ser Ile Ser Gly Thr Asp Gly Leu Ile Leu Leu Leu Val Gly Leu  
1 5 10 15

Glu Ala Xaa Val Arg Ser Ser Lys Lys Trp Ile Pro Lys Ala Leu Xaa  
20 25 30

Val Thr Gln Ala Lys Trp Asn Ser Trp Pro Ser Arg Arg Asn Ala Gly  
35 40 45

Phe Ala Leu His  
50

(2) INFORMATION FOR SEQ ID NO: 355:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 132 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

Met Glu His Cys Leu Tyr His Ser Val His Gly Ile Asn Pro Tyr Ile  
1 5 10 15

His Lys Asn Thr His Pro Ser Ile Asn Ile Tyr Met Val Trp Asp Glu  
20 25 30

Gln Val Asn Ser Phe Glu Arg Glu Phe Val Pro Phe Phe Leu Ile  
35 40 45

Ile Leu Leu Asn Cys Cys Gln Leu Ser Asn Lys Gln Thr Glu Lys Leu  
50 55 60

Phe Gly Lys Thr Leu His Thr Pro Phe Leu Ser Ser Ala Leu Lys Tyr  
65 70 75 80

Arg Leu Asn Thr His Ile Leu Pro Val Phe Ser Tyr Ser Asp Ser Ile  
85 90 95

Leu Thr Cys His Leu Ile Leu Ala Ser Tyr Phe Ser His Val Tyr Leu  
100 105 110

565

Pro Val Thr Cys Ile Cys Tyr Leu Asn Arg Lys Lys Asn Ile Gln Lys  
115 120

5 Lys Lys Asn Xaa  
130

10 (2) INFORMATION FOR SEQ ID NO: 356:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 204 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

Met Gly Ser Arg Asp His Leu Phe Lys Val Leu Val Val Gly Asp Ala  
1 5 10 15

20 Ala Val Gly Lys Thr Ser Leu Val Gln Asp Tyr Ser Gln Asp Ser Phe  
20 25 30

25 Ser Lys His Tyr Lys Ser Thr Val Gly Val Asp Phe Ala Leu Lys Val  
35 40 45

Leu Gln Trp Ser Asp Tyr Glu Ile Val Arg Leu Gln Leu Trp Asp Ile  
50 55 60

30 Ala Gly Gln Glu Arg Phe Thr Ser Met Thr Arg Leu Tyr Tyr Arg Asp  
65 70 75 80

Ala Ser Ala Cys Val Ile Met Phe Asp Val Thr Asn Ala Thr Thr Phe  
85 90 95

35 Ser Asn Ser Gln Arg Trp Lys Gln Asp Leu Asp Ser Lys Leu Thr Leu  
100 105 110

40 Pro Asn Gly Glu Pro Val Pro Cys Leu Leu Ala Asn Lys Cys Asp  
115 120 125

Leu Ser Pro Trp Ala Val Ser Arg Asp Gln Ile Asp Arg Phe Ser Lys  
130 135 140

45 Glu Asn Gly Phe Thr Gly Trp Thr Glu Thr Ser Val Lys Glu Asn Lys  
145 150 155 160

Asn Ile Asn Glu Ala Met Arg Val Leu Ile Glu Lys Met Met Arg Asn  
165 170 175

50 Ser Thr Glu Asp Ile Met Ser Leu Ser Thr Gln Gly Asp Tyr Ile Asn  
180 185 190

55 Leu Gln Thr Lys Ser Ser Trp Ser Cys Xaa  
195 200

(2) INFORMATION FOR SEQ ID NO: 357:

60

566

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

Met Ile Ser Leu Ile Phe Gln Leu Glu Glu Lys Leu Val Glu Lys  
1 5 10 15

10 Phe Phe Phe Leu Phe Phe Leu Lys Lys Gly Ser Gln Gly Ser  
20 25 30

Asn Leu Lys Ile Val Pro Arg His Met Arg Val Val Leu Arg Gly  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 358:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

Met Thr Tyr Val Thr Cys Leu His Val Cys Leu Leu Val Glu Phe Leu  
1 5 10 15

30 Asn Ser Gln Leu Thr Asn His Arg Lys Tyr Tyr Phe Leu Ser Tyr Gly  
20 25 30

Phe Trp Phe Thr Gly Leu Arg Gly Phe Ser Glu Tyr Tyr Leu Trp Pro Gln  
35 40 45

35 Gln His Thr Ser Phe His Pro Asn Arg Asn Glu Ile Asn Phe Val Ser  
50 55 60

Thr Asp Asn Arg Ile Trp Val Thr Xaa  
65 70

(2) INFORMATION FOR SEQ ID NO: 359:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 102 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

Met Ser Asp Gln Glu Ala Lys Pro Ser Thr Glu Asp Leu Gly Asp Lys  
1 5 10 15

55 Lys Glu Gly Glu Tyr Ile Lys Leu Lys Val Ile Gly Gln Asp Ser Ser  
20 25 30

Glu Ile His Phe Lys Val Lys Met Thr Thr His Leu Lys Lys Leu Lys  
35 40 45

60 Glu Ser Tyr Cys Gln Arg Gln Gly Val Pro Met Asn Ser Leu Arg Phe

50 55 60  
Leu Phe Glu Gly Gln Arg Ile Ala Asp Asn His Thr Pro Lys Glu Leu  
65 70 75 80  
Gly Met Glu Glu Glu Asp Val Ile Glu Val Tyr Gln Glu Gln Thr Gly  
85 90 95  
Gly His Ser Thr Val Xaa  
100

15 (2) INFORMATION FOR SEQ ID NO: 360:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

20 Met Gly Phe Pro Gln Trp His Leu Gly Asn His Ala Val Glu Pro Val  
1 5 10 15  
Thr Ser Ile Leu Leu Phe Leu Leu Met Met Leu Gly Val Arg Gly  
20 25 30  
Leu Leu Leu Val Gly Leu Val Tyr Leu Val Ser His Leu Ser Gln Arg  
35 40 45

30 (2) INFORMATION FOR SEQ ID NO: 361:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 179 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

40 Met Ser Ala Glu Val Lys Val Thr Gly Gln Asn Gln Glu Gln Phe Leu  
1 5 10 15  
Leu Leu Ala Lys Ser Ala Lys Gly Ala Ala Leu Ala Thr Leu Ile His  
20 25 30  
Gln Val Leu Glu Ala Pro Gly Val Tyr Val Phe Gly Glu Leu Leu Asp  
35 40 45  
Met Pro Asn Val Arg Glu Leu Ala Glu Ser Asp Phe Ala Ser Thr Phe  
50 55 60  
Arg Leu Leu Thr Val Phe Ala Tyr Gly Thr Tyr Ala Asp Tyr Leu Ala  
65 70 75 80  
Glu Ala Arg Asn Leu Pro Pro Leu Thr Glu Ala Gln Lys Asn Lys Leu  
85 90 95

Arg His Leu Ser Val Val Thr Leu Ala Ala Lys Val Lys Cys Ile Pro  
100 105 110  
Tyr Ala Val Leu Leu Glu Ala Leu Ala Leu Arg Asn Val Arg Gln Leu  
115 120 125  
Glu Asp Leu Val Ile Glu Ala Val Tyr Ala Asp Val Leu Arg Gly Ser  
130 135 140  
Leu Asp Gln Arg Asn Gln Arg Leu Glu Val Asp Tyr Ser Ile Gly Arg  
145 150 155 160  
Asp Ile Gln Arg Gln Asp Leu Ser Ala Ile Ala Arg Thr Leu Xaa Lys  
165 170 175  
Asn His Xaa

15 (2) INFORMATION FOR SEQ ID NO: 362:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

20 Met Lys Ser Ser Ser Leu Phe Phe Phe Leu Ala His Phe Ile His  
1 5 10 15  
Ser His Asp Leu Pro Gly Leu Cys Arg  
20 25

30 (2) INFORMATION FOR SEQ ID NO: 363:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 224 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

40 Met Lys Phe Ala Ala Ser Gly Xaa Phe Leu His His Met Ala Gly Leu  
1 5 10 15  
Ser Ser Ser Lys Leu Ser Met Ser Lys Ala Leu Pro Leu Thr Lys Val  
20 25 30  
Val Gln Asn Asp Ala Tyr Thr Ala Pro Ala Leu Pro Ser Ser Ile Arg  
35 40 45  
Thr Lys Ala Leu Thr Asn Met Ser Arg Thr Leu Val Asn Lys Glu Glu  
50 55 60  
Pro Pro Lys Glu Leu Pro Ala Ala Glu Pro Val Leu Ser Pro Leu Glu  
65 70 75 80

Gly Thr Lys Met Thr Val Asn Asn Leu His Pro Arg Val Thr Glu Glu 95  
85

5 Asp Ile Val Glu Leu Phe Cys Val Cys Gly Ala Leu Lys Arg Ala Arg 110  
100

Leu Val His Pro Gly Val Ala Glu Val Val Phe Val Lys Lys Asp Asp 125  
115

10 Ala Ile Thr Ala Tyr Lys Lys Tyr Asn Asn Arg Cys Leu Asp Gly Gln 140  
130

Pro Met Lys Cys Asn Leu His Met Asn Gly Asn Val Ile Thr Ser Asp 160  
145

15 Gln Pro Ile Leu Leu Arg Leu Ser Asp Ser Pro Ser Met Lys Lys Glu 175  
165

Ser Glu Leu Pro Arg Arg Val Asn Ser Ala Ser Ser Ser Asn Pro Pro 190  
180

Ala Glu Val Asp Pro Asp Thr Ile Leu Lys Ala Leu Phe Lys Ser Ser 205  
195

25 Gly Ala Ser Xaa Thr Thr Gln Pro Thr Glu Phe Lys Ile Lys Leu Xaa 220  
210

30

(2) INFORMATION FOR SEQ ID NO: 364:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 349 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

40 Met Ser Lys Asn Cys Ile Lys Leu Leu Cys Glu Asp Pro Val Phe Ala 15  
1

Glu Tyr Ile Lys Cys Ile Leu Met Asp Glu Arg Thr Phe Leu Asn Asn 30  
20

45 Asn Ile Val Tyr Thr Phe Met Thr His Phe Leu Leu Lys Val Gln Ser 45  
35

50 Gln Val Phe Ser Glu Ala Asn Cys Ala Asn Leu Ile Ser Thr Leu Ile 60  
50

Thr Asn Leu Ile Ser Gln Tyr Gln Asn Leu Gln Ser Asp Phe Ser Asn 80  
65

55 Arg Val Glu Ile Ser Lys Ala Ser Ala Ser Leu Asn Gly Asp Leu Arg 95  
85

Ala Leu Ala Leu Leu Ser Val His Thr Pro Lys Gln Leu Asn Pro 110  
100

Ala Leu Ile Pro Thr Leu Gln Glu Leu Ser Lys Cys Arg Thr Cys 125  
115

5 Leu Gln Gln Arg Asn Ser Leu Gln Glu Ala Lys Glu Arg Lys 140  
130

Thr Lys Asp Asp Glu Gly Ala Thr Pro Ile Lys Arg Arg Val Ser 155  
145

10 Ser Asp Glu Glu His Thr Val Asp Ser Cys Ile Ser Asp Met Lys Thr 175  
165

15 Glu Thr Arg Glu Val Leu Thr Pro Thr Ser Thr Ser Asp Asn Glu Thr 190  
180

Arg Asp Ser Ser Ile Ile Asp Pro Gly Thr Glu Gln Asp Leu Pro Ser 205  
195

20 Pro Glu Asn Ser Ser Val Lys Glu Tyr Arg Met Glu Val Pro Ser Ser 220  
210

25 Phe Ser Glu Asp Met Ser Asn Ile Arg Ser Gln His Ala Glu Glu Gln 240  
235

Ser Asn Asn Gly Arg Tyr Asp Cys Lys Glu Phe Lys Asp Leu His 255  
245

30 Cys Ser Lys Asp Ser Thr Leu Ala Glu Glu Ser Glu Phe Pro Ser 270  
260

Thr Ser Ile Ser Ala Val Leu Ser Asp Leu Ala Asp Leu Arg Ser Cys 285  
275

35 Asp Gly Gln Ala Leu Pro Ser Gln Asp Pro Glu Val Ala Leu Ser Leu 300  
290

Ser Cys Gly His Ser Arg Gly Leu Phe Ser His Met Gln Gln His Asp 315  
305

40 Ile Leu Asp Thr Leu Cys Arg Thr Ile Glu Ser Thr Ile His Val Val 335  
325

45 Thr Arg Ile Ser Gly Lys Gly Asn Gln Ala Ala Ser Xaa 345  
340

(2) INFORMATION FOR SEQ ID NO: 365:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 467 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 365:

50 Met Leu His Gln Asp His Ile Thr Phe Ala Met Leu Leu Ala Arg Ile 15  
1

55 Lys Leu Lys Gly Thr Val Gly Glu Pro Thr Tyr Asp Ala Glu Phe Gln 10

20 25 30  
 His Phe Leu Arg Gly Asn Glu Ile Val Leu Ser Ala Gly Ser Thr Pro  
 35 40 45  
 5 Arg Ile Glu Gly Leu Thr Val Glu Glu Ala Glu Ala Val Val Arg Leu  
 50 55 60  
 10 Ser Cys Leu Pro Ala Phe Lys Asp Leu Ile Ala Lys Val Glu Ala Asp  
 65 70 75 80  
 Glu Glu Phe Gly Ile Trp Leu Asp Ser Ser Pro Glu Glu Thr Val  
 85 90 95  
 15 Pro Tyr Leu Trp Ser Glu Glu Thr Pro Ala Thr Pro Ile Gly Glu Ala  
 100 105 110  
 Ile His Arg Leu Leu Leu Ile Glu Ala Phe Arg Pro Asp Arg Leu Leu  
 115 120 125  
 20 Ala Met Ala His Met Phe Val Ser Thr Asn Leu Leu Gly Glu Ser Phe Met  
 130 135 140  
 Ser Ile Met Glu Glu Pro Leu Asp Leu Thr His Ile Val Xaa Thr Glu  
 145 150 155 160  
 Val Lys Pro Asn Thr Pro Val Leu Met Cys Ser Val Pro Gly Tyr Asp  
 165 170 175  
 30 Ala Ser Gly His Val Glu Asp Leu Ala Ala Glu Glu Asn Thr Glu Ile  
 180 185 190  
 Thr Ser Ile Ala Ile Gly Ser Ala Glu Gly Phe Asn Glu Ala Asp Lys  
 195 200 205  
 35 Ala Ile Asn Thr Ala Val Lys Ser Gly Arg Trp Val Met Leu Lys Asn  
 210 215 220  
 Val His Leu Ala Pro Gly Trp Leu Met Glu Leu Glu Lys Lys Leu His  
 225 230 235 240  
 Ser Leu Glu Pro His Ala Cys Phe Arg Leu Phe Leu Thr Met Glu Ile  
 245 250 255  
 45 Asn Pro Lys Val Pro Val Asn Leu Leu Arg Ala Gly Arg Ile Phe Val  
 260 265 270  
 Phe Glu Pro Pro Gly Xaa Lys Ala Asn Met Leu Arg Thr Phe Ser  
 275 280 285  
 50 Ser Ile Pro Val Ser Arg Ile Cys Lys Ser Pro Asn Glu Arg Ala Arg  
 290 295 300  
 Leu Tyr Phe Leu Leu Ala Trp Phe His Ala Ile Ile Glu Glu Arg Leu  
 305 310 315 320  
 Arg Tyr Ala Pro Leu Gly Trp Ser Lys Tyr Glu Phe Gly Glu Ser  
 325 330 335  
 60 Asp Leu Arg Ser Xaa Cys Asp Thr Val Asp Thr Trp Leu Asp Asp Thr

340 345 350  
 Ala Lys Gly Arg Glu Asn Ile Ser Pro Asp Lys Ile Pro Trp Ser Ala  
 355 360 365  
 5 Leu Lys Thr Leu Met Ala Glu Ser Ile Tyr Gly Gly Arg Val Asp Asn  
 370 375 380  
 Glu Phe Asp Glu Arg Leu Leu Asn Thr Phe Leu Glu Arg Leu Phe Thr  
 385 390 395 400  
 Thr Arg Ser Phe Asp Ser Glu Phe Lys Leu Ala Cys Lys Val Asp Gly  
 405 410 415  
 15 His Lys Asp Ile Glu Met Pro Asp Gly Met Glu Ala Arg Gly Val Cys  
 420 425 430  
 Ala Val Gly Gly Val Ala Pro Arg His Pro Asp Ala Leu Leu Ala Gly  
 435 440 445  
 20 Pro Ala Glu Glu Arg Arg Glu Ser Pro Pro Tyr His Thr Gly Cys Gly  
 450 455 460  
 His Asp Glu  
 465  
 25  
 30 (2) INFORMATION FOR SEQ ID NO: 366:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 152 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 35 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 366:  
 Met Ala Asp Glu Ala Thr Arg Arg Val Val Ser Glu Ile Pro Val Leu  
 1 5 10 15  
 Lys Thr Asn Ala Gly Pro Arg Asp Arg Glu Leu Trp Val Glu Arg Leu  
 20 25 30  
 Lys Glu Glu Tyr Glu Ser Leu Ile Arg Tyr Val Glu Asn Asn Lys Asn  
 35 40 45  
 Ala Asp Asn Asp Trp Phe Arg Leu Glu Ser Asn Lys Glu Gly Thr Arg  
 50 55 60  
 Trp Phe Gly Lys Cys Trp Tyr Ile His Asp Leu Leu Lys Tyr Glu Phe  
 65 70 75 80  
 Asp Ile Glu Phe Asp Ile Pro Ile Thr Tyr Pro Thr Thr Ala Pro Glu  
 85 90 95  
 55 Ile Ala Val Pro Glu Leu Asp Gly Lys Thr Ala Lys Met Tyr Arg Gly  
 100 105 110  
 Gly Lys Ile Cys Leu Thr Asp His Phe Lys Pro Leu Trp Gly Glu Glu  
 115 120 125  
 60

Cys Ala Gln Ile Trp Thr Ser Ser His Gly Ser Gly Ala Gly Ser  
130 135 140

Met Xaa Gly Ser Gly Asn Pro Xaa  
145 150

10 (2) INFORMATION FOR SEQ ID NO: 367:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 373 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:

Met Tyr Asp Gly Thr Lys Glu Val Pro Met Asn Pro Val Lys Ile Tyr  
1 5 10 15

20 Gln Val Cys Asp Ile Pro Gln Pro Gln Gly Ser Ile Ile Asn Pro Gly  
20 25 30

25 Ser Thr Gly Ser Ala Pro Trp Asp Glu Lys Asp Asn Asp Val Asp Glu  
35 40 45

Glu Asp Glu Glu Asp Glu Leu Asp Gln Ser Gln His His Val Pro Ile  
50 55 60

30 Gln Asp Thr Phe Pro Phe Leu Asn Ile Asn Gly Ser Pro Met Ala Pro  
65 70 75 80

Ala Ser Val Gly Asn Cys Ser Val Gly Asn Cys Ser Pro Glu Ala Val  
85 90 95

35 Trp Pro Lys Thr Glu Pro Leu Glu Met Glu Val Pro Gln Ala Pro Ile  
100 105 110

40 Gln Pro Phe Tyr Ser Ser Pro Glu Leu Trp Ile Ser Ser Leu Pro Met  
115 120 125

Thr Asp Leu Asp Ile Lys Phe Gln Tyr Arg Gly Lys Glu Tyr Gly Gln  
130 135 140

45 Thr Met Thr Val Ser Asn Pro Gln Gly Cys Arg Leu Phe Tyr Gly Asp  
145 150 155 160

Leu Gly Pro Met Pro Asp Gln Glu Leu Phe Gly Pro Val Xaa Leu  
165 170 175

50 Glu Gln Val Lys Phe Pro Gly Pro Glu His Ile Thr Asn Glu Lys Gln  
180 185 190

Lys Leu Phe Thr Ser Lys Leu Leu Asp Val Met Asp Arg Gly Leu Ile  
195 200 205

55 Leu Glu Val Ser Gly His Ala Ile Tyr Ala Ile Arg Leu Cys Gln Cys  
210 215 220

60 Lys Val Tyr Trp Ser Gly Pro Cys Ala Pro Ser Leu Val Ala Pro Asn  
225 230 235 240

Leu Ile Glu Arg Gln Lys Lys Val Lys Leu Phe Cys Leu Glu Thr Phe  
245 250 255

5 Leu Ser Asp Leu Ile Ala His Gln Lys Gly Gln Ile Glu Lys Gln Pro  
260 265 270

10 Pro Phe Glu Ile Tyr Leu Cys Phe Gly Glu Glu Trp Pro Asp Gly Lys  
275 280 285

Pro Leu Glu Arg Lys Leu Ile Leu Val Gln Val Ile Pro Val Val Ala  
290 295 300

15 Arg Met Ile Tyr Glu Met Phe Ser Gly Asp Phe Thr Arg Ser Phe Asp  
305 310 315 320

Ser Gly Ser Val Arg Leu Gln Ile Ser Thr Pro Asp Ile Lys Asp Asn  
325 330 335

20 Ile Val Ala Gln Leu Lys Gln Leu Tyr Arg Ile Leu Gln Thr Gln Glu  
340 345 350

25 Ser Trp Gln Pro Met Gln Pro Thr Pro Ser Met Gln Leu Pro Pro Ala  
355 360 365

Leu Pro Pro Gln Xaa  
370

(2) INFORMATION FOR SEQ ID NO: 368:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser  
1 5 10 15

Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe  
20 25 30

45 Trp Gly Phe Gly Arg Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr  
35 40 45

Pro Pro Asp Gln Glu Pro Gln Lys Glu Thr Glu Pro Ala Thr Ser Ser  
50 55 60

His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln  
65 70 75 80

Pro Asn Xaa  
85

(2) INFORMATION FOR SEQ ID NO: 369:

60

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala  
1 5 10 15

10 Tyr Trp Thr Met Xaa  
20

15 (2) INFORMATION FOR SEQ ID NO: 370:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 227 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

Met Gly Ala Ser Ala Arg Leu Leu Arg Ala Val Ile Met Gly Ala Pro  
1 5 10 15

25 Gly Ser Gly Lys Gly Thr Val Ser Ser Arg Ile Thr Thr His Phe Glu  
20 25 30

30 Leu Lys His Leu Ser Ser Gly Asp Leu Leu Arg Asp Asn Met Leu Arg  
35 40 45

Gly Thr Glu Ile Gly Val Leu Ala Lys Ala Phe Ile Asp Glu Gly Lys  
50 55 60

35 Leu Ile Pro Asp Asp Val Met Thr Arg Leu Ala Leu His Glu Leu Lys  
65 70 75 80

Asn Leu Thr Glu Tyr Ser Trp Leu Leu Asp Gly Phe Pro Arg Thr Leu  
85 90 95

40 Pro Glu Ala Glu Ala Leu Asp Arg Ala Tyr Glu Ile Asp Thr Val Ile  
100 105 110

45 Asn Leu Asn Val Pro Phe Glu Val Ile Lys Glu Arg Leu Thr Ala Arg  
115 120 125

Trp Ile His Pro Ala Ser Gly Arg Val Tyr Asn Ile Glu Phe Asn Pro  
130 135 140

50 Pro Lys Thr Val Gly Ile Asp Asp Leu Thr Gly Glu Pro Leu Ile Glu  
145 150 155 160

Arg Glu Asp Asp Lys Pro Glu Thr Val Ile Lys Arg Leu Lys Ala Tyr  
165 170 175

55 Glu Asp Glu Thr Lys Pro Val Leu Glu Tyr Tyr Glu Lys Lys Val  
180 185 190

60 Leu Glu Thr Phe Ser Gly Thr Glu Thr Asn Lys Ile Trp Pro Tyr Val  
195 200 205

Tyr Ala Phe Leu Glu Thr Lys Val Pro Glu Arg Ser Glu Lys Ala Ser  
210 215 220

5 Val Thr Pro  
225

10 (2) INFORMATION FOR SEQ ID NO: 371:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

Met Phe Leu Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Glu  
1 5 10 15

20 Leu Lys Ile Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu  
20 25 30

25 Lys Ile Ile Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Ser  
35 40 45

Trp Ala Ile Lys Ala Glu Leu Lys Ile Glu Asn Lys Asp Leu Asp Asn  
50 55 60

30 Lys Thr Ala Lys Gly Gly Glu Glu Ala Leu Thr Cys Thr Xaa  
65 70 75

35 (2) INFORMATION FOR SEQ ID NO: 372:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

Met Arg Ala Val Phe Pro Cys Cys Pro Phe Leu Thr Leu Met Leu Pro  
1 5 10 15

45 Leu Leu Glu Cys Leu Val Gly Met Ile Met Cys Tyr Leu Gly Ile Ser  
20 25 30

50 Phe Thr Asp Thr Arg Lys Thr Ala Gly Leu Lys Lys Lys Lys  
35 40 45

Lys Xaa Xaa  
50

55 (2) INFORMATION FOR SEQ ID NO: 373:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids



577

578

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

5 Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser 15  
1  
Phe Ile Val Thr Pro Thr Ser Thr Ser Asn Thr Ala Ser Tyr Leu Trp 30  
20  
10 Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Ser Trp Ala Cys 45  
35  
Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu Xaa 60  
50

(2) INFORMATION FOR SEQ ID NO: 374:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

25 Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe 15  
1  
30 Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg 30  
20  
Ile Leu Phe Phe Ile Val Phe Xaa 40  
35

(2) INFORMATION FOR SEQ ID NO: 375:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 44 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

45 Met Cys Ser Gly Gln Ser Gln Val Trp Lys Met Ala Leu Gln Ala Leu 15  
1  
40 Asp Ser Glu Thr Val Val Ile Leu Pro Asp Met His Leu Ile Leu Ser 30  
20  
50 Leu Arg Leu Ile His Asn Ala Arg Pro Cys Leu Xaa 40  
35

(2) INFORMATION FOR SEQ ID NO: 376:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 203 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

5 Met Leu Ile Ser Glu Glu Ile Pro Phe Lys Asp Asp Pro Arg Asp 15  
1  
10 Glu Thr Tyr Lys Pro His Leu Glu Arg Glu Thr Pro Lys Pro Arg Arg 30  
20  
15 Lys Ser Gly Lys Val Lys Glu Glu Lys Glu Lys Lys Glu Ile Lys Val 45  
35  
Glu Val Glu Val Glu Val Lys Glu Glu Glu Asn Glu Ile Arg Glu Asp 60  
50  
Glu Glu Pro Pro Arg Lys Arg Gly Arg Arg Arg Lys Asp Lys Ser 80  
65  
20 Pro Arg Leu Pro Lys Arg Lys Lys Pro Pro Ile Gln Tyr Val Arg 95  
85  
Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His Pro Arg Tyr Leu 110  
100  
25 Gln His His Ile Lys Tyr Gln His Leu Leu Lys Lys Tyr Val Cys 125  
115  
30 Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln Lys Gln Leu Leu 140  
130  
Arg His Ala Lys His Thr Asp Gln Arg Asp Tyr Ile Cys Glu Tyr 160  
145  
35 Cys Ala Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met 175  
165  
Ile His Thr Gly Glu Lys His Tyr Asn Val Arg Ser Val Asp Leu Leu 190  
180  
40 Val Asp Lys Arg His Leu Leu Ile Gly Thr Xaa 200  
195  
45 (2) INFORMATION FOR SEQ ID NO: 377:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

55 Met Leu Pro Arg Arg Thr Phe Tyr Phe Thr Phe Ile Phe Phe Phe 15  
1  
Leu Ala Ser Phe Trp Gly Phe Thr Leu Arg Ala Ser Phe 25  
20

579

## (2) INFORMATION FOR SEQ ID NO: 378:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 116 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

10 Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Met  
1 5 10 15  
Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val Leu  
20 25 30  
Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu Trp  
35 40 45  
Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg Met  
50 55 60  
Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val Gln  
65 70 75 80  
Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu Asn  
85 90 95  
Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met Pro Ser Phe  
100 105 110  
Phe Ser Cys His Leu Phe Cys Thr Leu Arg Trp Lys Tyr Phe Glu Val  
115 120 125  
Phe Tyr Asn His Lys Phe Leu Xaa  
130 135

## (2) INFORMATION FOR SEQ ID NO: 379:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

45 Met Ala Trp Arg Arg Arg Glu Pro Ala Ser Gly Leu Ala Ala Cys Trp  
1 5 10 15  
Leu Trp Arg Cys Ser Pro Trp Pro Cys Ala Cys Pro Gly Pro Gly Ala  
20 25 30  
Gly Leu Ser Ser Gly Ser Arg Pro Trp  
35 40

## (2) INFORMATION FOR SEQ ID NO: 380:

## (1) SEQUENCE CHARACTERISTICS:

580

(A) LENGTH: 468 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

5 Met Glu Phe Leu Lys Val Ala Arg Arg Asn Lys Arg Glu Gln Leu Glu  
1 5 10 15  
Gln Ile Gln Lys Glu Leu Ser Val Leu Glu Glu Asp Ile Lys Arg Val  
20 25 30  
Glu Glu Met Ser Gly Leu Tyr Ser Pro Val Ser Glu Asp Ser Thr Val  
35 40 45  
Pro Gln Phe Glu Ala Pro Ser Pro Ser His Ser Ser Ile Ile Asp Ser  
50 55 60  
Thr Glu Tyr Ser Gln Pro Pro Gly Phe Ser Gly Ser Ser Gln Thr Lys  
65 70 75 80  
Lys Gln Pro Trp Tyr Asn Ser Thr Leu Ala Ser Arg Arg Lys Arg Leu  
85 90 95  
Thr Ala His Phe Glu Asp Leu Glu Gln Cys Tyr Phe Ser Thr Arg Met  
100 105 110  
Ser Arg Ile Ser Asp Asp Ser Arg Thr Ala Ser Gln Leu Asp Glu Phe  
115 120 125  
Gln Glu Cys Leu Ser Lys Phe Thr Arg Tyr Asn Ser Val Arg Pro Leu  
130 135 140  
Ala Thr Leu Ser Tyr Ala Ser Asp Leu Tyr Asn Gly Ser Ser Ile Val  
145 150 155 160  
Ser Ser Ile Glu Phe Asp Arg Asp Cys Asp Tyr Phe Ala Ile Ala Gly  
165 170 175  
Val Thr Lys Lys Ile Lys Val Tyr Glu Tyr Asp Thr Val Ile Gln Asp  
180 185 190  
Ala Val Asp Ile His Tyr Pro Glu Asn Glu Met Thr Cys Asn Ser Lys  
195 200 205  
Ile Ser Cys Ile Ser Trp Ser Ser Tyr His Lys Asn Leu Leu Ala Ser  
210 215 220  
Ser Asp Tyr Glu Gly Thr Val Ile Leu Trp Asp Gly Phe Thr Gly Gln  
225 230 235 240  
Arg Ser Lys Val Tyr Gln Glu His Glu Lys Arg Cys Trp Ser Val Asp  
245 250 255  
Phe Asn Leu Met Asp Pro Lys Leu Leu Ala Ser Gly Ser Asp Asp Ala  
260 265 270  
Lys Val Lys Leu Trp Ser Thr Asn Leu Asp Asn Ser Val Ala Ser Ile  
275 280 285  
Glu Ala Lys Ala Asn Val Cys Cys Val Lys Phe Ser Pro Ser Ser Arg  
290 295 300

582

Met Pro Leu Ala Pro Tyr Cys Asp Leu Leu Val Ala Leu Ser Phe Ala  
1 5 10 15  
5 Leu Val Leu Glu Ser Pro Val Asp Ser Ser Asp Phe Thr  
20 25

10 (2) INFORMATION FOR SEQ ID NO: 383:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 138 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

Met Asn Ser Leu Val Ser Trp Gln Leu Leu Phe Leu Cys Ala Thr  
1 5 10 15  
20  
His Phe Gly Glu Pro Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg  
20 25 30

Pro Thr Gly Gln Gln Leu Glu Ser Leu Gly Leu Ala Pro Gly Glu  
25 35 40 45  
Gln Ser Leu Pro Cys Thr Glu Arg Lys Pro Ala Thr Ala Arg Leu  
50 55 60

25 Asp Thr Val Lys Ser Val Leu Asp Lys Asp Arg Lys Glu Asp Asp Thr 420 425 430

Asn Glu Phe Val Ser Ala Val Cys Trp Arg Ala Leu Pro Asp Gly Glu 435 440 445

30 Ser Arg Arg Gly Thr Ser Leu Ser Pro Pro Glu Ser Ser Gly Ser  
65 70 75 80  
Pro Gln Gln Pro Gly Leu Ser Ala Pro His Ser Arg Gln Ile Pro Ala  
95

35 Pro Gln Gly Ala Val Leu Val Gln Arg Glu Lys Asp Leu Pro Asn Tyr  
100 105 110

40 Asn Trp Asn Ser Phe Gly Leu Arg Phe Gly Lys Arg Glu Ala Ala Pro  
115 120 125

Gly Asn His Gly Arg Ser Ala Gly Arg Gly  
130 135

(2) INFORMATION FOR SEQ ID NO: 384:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(vi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

55 Met Ser Cys Phe Ile Asp Ser Xaa Asp Ser Lys Ile Leu His Leu Leu 15  
1 5 10  
Val Val Ser Phe Ile Cys Xaa Leu Phe Leu Leu Ile Leu Thr His Gly 30  
20 25

Ile Leu Ile Leu Arg Xaa Phe Ser Val Xaa Xaa His Ser Leu Lys  
35 40 45  
Asn Asn Leu Glu Glu Tyr Leu Ile Leu Met Asn Lys Ala Leu Leu Thr  
5 50 55 60  
Arg Glu Asp Phe Phe Val Leu Pro Xaa Ala  
65 70  
(2) INFORMATION FOR SEQ ID NO: 385:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 521 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 385:  
Met Ser Ala Gly Glu Val Glu Arg Leu Val Ser Glu Leu Ser Gly Gly  
1 5 10  
Thr Gly Gly Asp Glu Glu Glu Glu Tyr Leu Tyr Gly Asp Glu Asn Glu  
20 25 30  
Val Glu Arg Pro Glu Glu Glu Asn Ala Ser Ala Asn Pro Ser Gly  
35 40 45  
Ile Glu Asp Glu Thr Ala Glu Asn Gly Val Pro Lys Pro Lys Val Thr  
50 55 60  
Glu Thr Glu Asp Asp Ser Asp Ser Asp Asp Asp Glu Asp Asp  
65 70 75 80  
Val His Val Thr Ile Gly Asp Ile Lys Thr Gly Ala Pro Glu Tyr Gly  
85 90 95  
Ser Tyr Gly Thr Ala Pro Val Asn Leu Asn Ile Lys Thr Gly Gly Arg  
100 105 110  
Val Tyr Gly Thr Thr Gly Thr Lys Val Lys Gly Val Asp Leu Asp Ala  
115 120 125  
Pro Gly Ser Ile Asn Gly Val Pro Leu Leu Glu Val Asp Leu Asp Ser  
130 135 140  
Phe Glu Asp Lys Pro Tyr Arg Lys Pro Gly Ala Asp Leu Ser Asp Tyr  
145 150 155 160  
Phe Asn Tyr Gly Phe Asn Glu Asp Thr Tyr Lys Ala Tyr Cys Glu Lys  
165 170 175  
Gln Lys Arg Ile Arg Met Gly Leu Glu Val Ile Pro Val Thr Ser Thr  
180 185 190  
Thr Asn Lys Ile Thr Val Gln Gln Gly Arg Thr Gly Asn Ser Glu Lys  
195 200 205  
Glu Thr Ala Leu Pro Ser Thr Lys Ala Glu Phe Thr Ser Pro Pro Ser  
210 215 220

Leu Phe Lys Thr Gly Leu Pro Pro Ser Arg Arg Leu Pro Gly Ala Ile  
225 230 235 240  
Asp Val Ile Gly Gln Thr Ile Thr Ile Ser Arg Val Glu Gly Arg Arg  
245 250 255  
Arg Ala Asn Glu Asn Ser Asn Ile Gln Val Leu Ser Glu Arg Ser Ala  
260 265 270  
Thr Glu Val Asp Asn Asn Phe Ser Lys Pro Pro Phe Phe Pro  
275 280 285  
Gly Ala Pro Thr His Leu Pro Pro Pro Pro Phe Leu Pro Pro  
290 295 300  
Pro Thr Val Ser Thr Ala Pro Pro Leu Ile Pro Pro Gly Phe Pro  
305 310 315 320  
Pro Pro Gly Ala Pro Pro Ser Ser Leu Ile Pro Thr Ile Glu Ser  
325 330 335  
Gly His Ser Ser Gly Tyr Asp Ser Arg Ser Ala Arg Ala Phe Pro Tyr  
340 345 350  
Gly Asn Val Ala Phe Pro His Leu Pro Gly Ser Ala Pro Ser Thr Pro  
355 360 365  
Ser Leu Val Asp Thr Ser Lys Gln Tyr Asp Tyr Tyr Ala Arg Arg Glu  
370 375 380  
Lys Asp Arg Asp Arg Glu Arg Asp Arg Asp Glu Arg Asp Arg Asp  
385 390 395 400  
Arg Asp Arg Glu Arg Glu Arg Thr Arg Glu Arg Glu Arg Glu Arg Asp  
405 410 415  
His Ser Pro Thr Pro Ser Val Phe Asn Ser Asp Glu Glu Arg Tyr Arg  
420 425 430  
Tyr Arg Glu Tyr Ala Glu Arg Gly Tyr Glu Arg His Arg Ala Ser Arg  
435 440 445  
Glu Lys Glu Glu Arg His Arg Glu Arg Arg His Arg Glu Lys Glu Glu  
450 455 460  
Thr Arg His Lys Ser Ser Arg Ser Asn Ser Arg Arg His Glu Ser  
465 470 475 480  
Glu Glu Gly Asp Ser His Arg Arg His Lys Lys Lys Ser Lys Arg  
485 490 495  
Ser Lys Glu Gly Lys Glu Ala Gly Ser Glu Pro Ala Pro Glu Gln Glu  
500 505 510  
Ser Thr Glu Ala Thr Pro Ala Glu Xaa  
515 520 525

## (2) INFORMATION FOR SEQ ID NO: 386:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 137 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 386:

Met Asn Ser Arg Gly Ile Trp Leu Ala Tyr Ile Ile Leu Val Gly Leu 15  
 1 5 10  
 Leu His Met Val Leu Leu Ser Ile Pro Phe Phe Ser Ile Pro Val Val 20  
 20 25 30  
 Trp Thr Leu Thr Asn Val Ile His Asn Leu Ala Thr Tyr Val Phe Leu 35  
 40 45  
 His Thr Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala 50  
 55 60  
 Arg Leu Leu Thr His Trp Glu Gln Met Asp Tyr Gly Leu Gln Phe Thr 65  
 70 75 80  
 Ser Ser Arg Lys Phe Leu Ser Ile Ser Pro Ile Val Leu Tyr Leu Leu 85  
 90 95  
 Ala Ser Phe Tyr Thr Lys Tyr Asp Ala Ala His Phe Leu Ile Asn Thr 100  
 105 110  
 Ala Ser Leu Leu Ser Val Leu Leu Pro Lys Leu Pro Gln Phe His Gly 115  
 120 125  
 Val Arg Val Phe Gly Ile Asn Lys Tyr 130  
 135

## (2) INFORMATION FOR SEQ ID NO: 387:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 186 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 387:

Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu 15  
 1 5 10  
 Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg 20  
 25 30  
 Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe 35  
 40 45  
 Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Leu Cys 50  
 55 60  
 Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe 65  
 70 75 80

Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu 95  
 85 90  
 Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Tyr Ile Leu Tyr Leu 100  
 105 110  
 Arg Thr Leu Glu Ser Lys Leu Val Leu Phe Gly Arg Glu Val Ser Pro 115  
 120 125  
 Ala His Gln Tyr Ala Leu Ala Gly Ile Ser Phe Pro Phe Thr Trp 130  
 135 140  
 Leu Ala Gly Ala Gly Ser Ala Val Phe Trp Val Leu Gly Ala Thr Leu 145  
 150 155 160  
 Val Val Ile Gly Ser His Ala Ala Phe His Gln Ile Glu Ala Val Asp 165  
 170 175  
 Gly Glu Glu Leu Gln Met Glu Pro Val Xaa 180  
 185

## (2) INFORMATION FOR SEQ ID NO: 388:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

Met 1

## (2) INFORMATION FOR SEQ ID NO: 389:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 299 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

Met Leu Ser Ile Phe Tyr Phe Ala Ile Pro Val Gly Ser Gly Leu Gly 15  
 1 5 10  
 Tyr Ile Ala Gly Ser Lys Val Lys Asp Met Ala Gly Asp Trp His Trp 20  
 25 30  
 Ala Leu Arg Val Thr Pro Gly Leu Gly Val Val Ala Val Leu Leu Leu 35  
 40 45  
 Phe Leu Val Val Arg Glu Pro Pro Arg Gly Ala Val Glu His Ser 50  
 55 60  
 Asp Leu Pro Pro Leu Asn Pro Thr Ser Trp Ala Asp Leu Arg Ala 65  
 70 75 80  
 Leu Ala Arg Asn Pro Ser Phe Val Leu Ser Ser Leu Gly Phe Thr Ala 90

85 90 95

Val Ala Phe Val Thr Gly Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu  
100 105 110

5 Leu Arg Ser Arg Val Val Leu Gly Glu Thr Pro Cys Leu Pro Gly  
115 120 125

10 Asp Ser Cys Ser Ser Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys  
130 135 140

Leu Thr Gly Val Leu Gly Val Gly Leu Gly Val Glu Ile Ser Arg Arg  
145 150 155 160

15 Leu Arg His Ser Asn Pro Arg Ala Asp Pro Leu Val Cys Ala Thr Gly  
165 170 175

Leu Leu Gly Ser Ala Pro Phe Leu Phe Leu Ser Leu Ala Cys Ala Arg  
180 185 190

20 Gly Ser Ile Val Ala Thr Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu  
195 200 205

25 Leu Ser Met Asn Trp Ala Ile Val Ala Asp Ile Leu Leu Tyr Val Val  
210 215 220

Ile Pro Thr Arg Arg Ser Thr Ala Glu Ala Phe Glu Ile Val Leu Ser  
225 230 235 240

30 His Leu Leu Gly Asp Ala Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser  
245 250 255

Asp Arg Leu Arg Arg Asn Trp Pro Pro Ser Phe Leu Ser Glu Phe Arg  
260 265 270

35 Ala Leu Glu Phe Ser Leu Met Leu Cys Ala Phe Val Gly Ala Leu Gly  
275 280 285

40 Gly Ala Leu Pro Gly His Arg His Leu His Xaa  
290 295

(2) INFORMATION FOR SEQ ID NO: 390:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 49 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

Met Gly Pro Glu Gly Trp Val Arg Pro Leu Lys Thr Ala Pro Lys Leu  
1 5 10 15

55 Gly Glu Ala Ile Arg Leu Ile Leu Phe Leu Asn Phe Val Lys Glu Cys  
20 25 30

Ile Ala Ser Val Asn Leu Cys Ile Leu Arg Leu Asn Ile Thr Pro Leu  
35 40 45

Leu

5 (2) INFORMATION FOR SEQ ID NO: 391:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 391:

Met Tyr Val Asn Tyr Gly Thr Arg Asn Tyr Ser Thr Glu Gly Pro Ala  
1 5 10 15

Ala Leu Leu Asp Glu Ala Lys Leu Ser Leu Leu Val Trp Val Leu Cys  
20 25 30

20 Phe Val Leu Leu Phe Val Cys Phe Cys Gly Leu Ser Tyr Val Val Ile  
35 40 45

25 Ala Glu Val Pro Val Gly Leu Leu Cys Ile Thr Glu Xaa  
50 55 60

(2) INFORMATION FOR SEQ ID NO: 392:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 79 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 392:

Met Leu Trp Phe Ala Asn Phe Phe Thr Tyr Leu Phe Leu Ser Glu Ser  
1 5 10 15

Val Ala Phe Val His Ile Ser His Ile Gly Val Arg Glu Val Asn Thr  
20 25 30

Asn Cys Tyr Phe Ser Arg Lys Ser Tyr Cys Tyr Gly Ile Leu Asn Pro  
35 40 45

45 Ile Asn Cys Ile Lys Gly Lys Lys Lys Lys Lys Lys Lys Lys Lys  
50 55 60

Lys Lys Lys Lys Ile Pro Ala Gly Arg Xaa Leu Phe Pro Phe Gly  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 393:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 393:

589

Met Pro Gly Ala Phe Ser Glu Thr Val Ile Asn Asp Leu Leu Ser Leu  
1 5 10 15

Phe Leu Val Leu Pro Ala Glu Leu Ser Tyr Ser Thr Leu Ser Gly Val  
5 20 25 30

Tyr Arg Asn Ala  
35

10

(2) INFORMATION FOR SEQ ID NO: 394:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 180 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 394:

Met Ala Gln Ser Arg Asp Gly Gly Asn Pro Phe Ala Glu Pro Ser Glu  
1 5 10 15

Leu Asp Asn Pro Phe Gln Asp Pro Ala Val Ile Gln His Arg Pro Ser  
20 25 30

Arg Gln Tyr Ala Thr Leu Asp Val Tyr Asn Pro Phe Glu Thr Arg Glu  
35 40 45

Pro Pro Pro Ala Tyr Glu Pro Pro Ala Pro Ala Pro Leu Pro Pro Pro  
50 55 60

Ser Ala Pro Ser Leu Gln Pro Ser Arg Lys Leu Ser Pro Thr Glu Pro  
65 70 75 80

Lys Asn Tyr Gly Ser Tyr Ser Thr Gln Ala Ser Ala Ala Ala Thr  
85 90 95

Ala Glu Leu Leu Lys Lys Gln Glu Glu Leu Asn Arg Lys Ala Glu Glu  
100 105 110

Leu Asp Arg Arg Ser Glu Ser Cys Ser Met Leu Pro Thr Xaa Ala Gln  
115 120 125

Leu Leu Asp Arg Thr Ile Gly Pro Leu Tyr Leu Phe Val Gln Phe  
130 135 140

Ser Pro Ala Phe Ser Arg Thr Ser Pro Thr Arg Ser Pro Lys Asn Phe  
145 150 155 160

Arg Arg Leu Tyr Pro Pro Cys Thr Thr Ser Gly Cys Ala Ala Arg Thr  
165 170 175

Xaa Phe Ser Xaa  
180

55

(2) INFORMATION FOR SEQ ID NO: 395:

(1) SEQUENCE CHARACTERISTICS:

590

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 395:

Met Pro Thr Pro Cys Thr Ser Leu Pro Ser Cys Cys Gln His Arg Ser  
1 5 10 15

Ile Thr Met Thr Leu  
20

10

(2) INFORMATION FOR SEQ ID NO: 396:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 396:

Met Pro Leu Phe Ile Pro Leu Ile Phe Leu Ser Leu Leu His Cys  
1 5 10 15

Gln Ser Lys His Pro Ile Gln Met Ser Leu Cys Met Cys Val Asn Ile  
20 25 30

Ser Leu Val Trp Ser Pro Val Arg Trp Ile Phe Gly Ser Lys Gly Leu  
35 40 45

Phe Ser Val His Leu Gln Ser Ser Gln Arg Pro Ser  
50 55 60

35

(2) INFORMATION FOR SEQ ID NO: 397:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 152 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 397:

Met Ala Gly Pro Arg Pro Xaa Trp Arg Asp Gln Leu Leu Phe Met Ser  
1 5 10 15

Ile Ile Val Leu Val Ile Val Ile Cys Leu Met Leu Tyr Ala Leu  
20 25 30

Leu Trp Glu Ala Gly Asn Leu Thr Asp Leu Pro Asn Leu Arg Ile Gly  
35 40 45

Phe Tyr Asn Phe Cys Leu Trp Asn Glu Asp Thr Ser Thr Leu Gln Cys  
50 55 60

His Gln Phe Pro Glu Leu Glu Ala Leu Gly Val Pro Arg Val Gly Leu  
65 70 75 80

Gly Leu Ala Arg Leu Gly Val Tyr Gly Ser Leu Val Leu Thr Leu Phe  
85 90 95

60

Ala Pro Gln Pro Leu Leu Leu Ala Gln Cys Asn Xaa Asp Gln Arg Ala  
100 105 110  
5 Trp Arg Leu Ala Val Gly Phe Leu Ala Val Ser Ser Val Leu Leu Ala  
115 120 125  
10 Gly Gly Leu Gly Leu Phe Leu Ser Tyr Val Trp Asn Gly Ser Xaa Ser  
130 135 140  
Pro Ser Arg Gly Leu Gly Phe Xaa  
145 150  
15 (2) INFORMATION FOR SEQ ID NO: 398:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 480 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 398:  
25 Met Ser Asp Gly Phe Asp Arg Ala Pro Gly Ala Gly Arg Gly Arg Xaa  
1 5 10 15  
Arg Gly Leu Gly Arg Gly Gly Gly Pro Xaa Gly Gly Gly Phe Pro  
20 25 30  
30 Xaa Gly Xaa Xaa Pro Ala Gln Arg Xaa Arg His Gln Pro Pro Gln Pro  
35 40 45  
Lys Ala Pro Gly Phe Leu Gln Pro Xaa Pro Leu Arg Gln Pro Arg Thr  
50 55 60  
35 Thr Pro Pro Gly Ala Gln Cys Gln Val Pro Ala Ser Pro Gln Arg  
65 70 75 80  
40 Pro Ser Arg Pro Gly Ala Leu Pro Gln Gln Thr Arg Pro Leu Arg Ala  
85 90 95  
Pro Pro Ser Ser Gln Asp Lys Ile Pro Gln Gln Asn Ser Gln Ser Ala  
100 105 110  
45 Met Ala Lys Pro Gln Val Val Ala Pro Val Leu Met Ser Lys Leu  
115 120 125  
Ser Val Asn Ala Pro Gln Phe Tyr Pro Ser Gly Tyr Ser Ser Ser Tyr  
130 135 140  
50 Thr Gln Ser Tyr Gln Asp Gly Cys Gln Asp Tyr Pro Thr Leu Ser Gln  
145 150 155 160  
55 Tyr Val Gln Asp Phe Leu Asn His Leu Thr Gln Gln Pro Gly Ser Phe  
165 170 175  
Glu Thr Gln Ile Gln Gln Phe Ala Gln Thr Leu Asn Gly Cys Val Thr  
180 185 190  
60 Thr Asp Asp Ala Leu Gln Gln Leu Val Gln Leu Ile Tyr Gln Gln Ala

195 200 205  
5 Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr  
210 215 220  
Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln  
225 230 235 240  
10 Leu Leu Leu Gln Arg Cys Arg Thr Gln Tyr Gln Val Lys Asp Gln Ala  
245 250 255  
Ala Lys Gly Asp Gln Val Thr Arg Lys Arg Phe His Ala Phe Val Leu  
260 265 270  
15 Phe Leu Gly Gln Leu Tyr Leu Asn Leu Gln Ile Lys Gly Thr Asn Gly  
275 280 285  
Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Gln Leu Leu  
290 295 300  
20 Asn Ala Leu Phe Ser Asn Pro Met Asp Asp Asn Leu Ile Cys Ala Val  
305 310 315 320  
Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Gln Asp Ala Trp Lys Gln  
325 330 335  
25 Lys Gly Lys Met Asp Met Gln Gln Ile Ile Gln Arg Ile Gln Asn Val  
340 345 350  
30 Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Lys  
355 360 365  
Leu Val Gln Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser  
370 375 380  
35 Thr Tyr Arg Gln Ala Thr Gln Asn Asp Pro Asn Tyr Phe Met Asn  
385 390 395 400  
40 Gln Pro Thr Phe Tyr Thr Ser Asp Gly Val Phe Thr Ala Ala Asp  
405 410 415  
Pro Asp Tyr Gln Gln Lys Tyr Gln Gln Leu Leu Gln Arg Gln Asp Phe  
420 425 430  
45 Phe Pro Asp Tyr Gln Gln Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp  
435 440 445  
Pro Tyr Leu Asp Asp Ile Asp Asp Gln Met Asp Pro Gln Ile Gln Gln  
450 455 460  
50 Ala Tyr Gln Lys Phe Cys Leu Gln Ser Gln Arg Lys Arg Lys Gln Xaa  
465 470 475 480  
55 (2) INFORMATION FOR SEQ ID NO: 399:  
60



WO 98/39448

594

Ile Thr Leu Met Gly Pro Gln Trp Trp Leu Lys Thr Val Ile Glu Gln  
290 295 300  
Val Tyr Ala Asn Gly Ile Arg Asn Ile Asp Leu His Tyr Ile Val Arg  
305 310 315 320  
Lys Leu Ala Ala Pro Val Ile Ser Val Leu Leu Ser Leu Cys Val  
325 330 335  
10 Pro Tyr Val Ile Ala Ser Gly Val Val Pro Leu Leu Gly Val Thr Ala  
340 345 350  
Glu Met Gln Asn Leu Val His Arg Arg Ile Tyr Pro Phe Leu Leu Met  
355 360 365  
15 Val Val Val Leu Met Ala Ile Leu Ser Phe Gln Val Arg Gln Phe Lys  
370 375 380  
Arg Leu Tyr Glu His Ile Lys Asn Asp Lys Tyr Leu Val Gly Gln Arg  
385 390 395 400  
Leu Val Asn Tyr Glu Arg Lys Ser Gly Lys Gln Gly Ser Ser Pro Pro  
405 410 415  
25 Pro Pro Gln Ser Ser Gln Glu  
420

(1) INFORMATION FOR SEQ ID NO: 400:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 400:

Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met  
1 5 10 15  
Leu Ile Val Ser Val Leu Ala Leu Ile Pro Glu Thr Thr Thr Leu Thr  
20 25 30  
Val Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Cys Leu Ala  
35 40 45  
Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro  
50 55 60  
50 Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu Xaa  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 401:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

PCT/US98/04493

WO 98/39448

593

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 423 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 399:

Met Glu Pro Lys Thr Ile Thr Asp Ala Leu Ala Ser Ser Ile Ile Lys  
1 5 10 15  
10 Ser Val Leu Pro Asn Phe Leu Pro Tyr Asn Val Met Leu Tyr Ser Asp  
20 25 30  
Ala Pro Val Ser Glu Leu Ser Leu Glu Leu Leu Leu Gln Val Val  
35 40 45  
15 Leu Pro Ala Leu Leu Glu Gln Gly His Thr Arg Gln Trp Leu Lys Gly  
50 55 60  
Leu Val Arg Ala Trp Thr Val Thr Ala Gly Tyr Leu Leu Asp Leu His  
65 70 75 80  
20 Ser Tyr Leu Leu Gly Asp Gln Glu Glu Asn Glu Asn Ser Ala Asn Gln  
85 90 95  
25 Gln Val Asn Asn Asn Gln His Ala Arg Asn Asn Asn Ala Ile Pro Val  
100 105 110  
Val Gly Glu Gly Leu His Ala Ala His Gln Ala Ile Leu Gln Gln Gly  
115 120 125  
30 Gly Pro Val Gly Phe Gln Xaa Tyr Arg Arg Pro Leu Asn Phe Pro Leu  
130 135 140  
Arg Ile Phe Leu Leu Ile Val Phe Met Cys Ile Thr Leu Leu Ile Ala  
145 150 155 160  
35 Ser Leu Ile Cys Leu Thr Leu Pro Val Phe Ala Gly Arg Trp Leu Met  
165 170 175  
40 Ser Phe Trp Thr Gly Thr Ala Lys Ile His Glu Leu Tyr Thr Ala Ala  
180 185 190  
Cys Gly Leu Tyr Val Cys Trp Leu Thr Ile Arg Ala Val Thr Val Met  
195 200 205  
45 Val Ala Trp Met Pro Gln Gly Arg Arg Val Ile Phe Gln Lys Val Lys  
210 215 220  
Glu Trp Ser Leu Met Ile Met Lys Thr Leu Ile Val Ala Val Leu Leu  
225 230 235 240  
50 Ala Gly Val Val Pro Leu Leu Leu Gly Leu Leu Phe Glu Leu Val Ile  
245 250 255  
55 Val Ala Pro Leu Arg Val Pro Leu Asp Gln Thr Pro Leu Phe Tyr Pro  
260 265 270  
Trp Gln Asp Trp Ala Leu Gly Val Leu His Ala Lys Ile Ile Ala Ala  
275 280 285  
60

595

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 401:

Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser  
1 5 10 15  
His Cys Trp Gly Leu Pro Leu His Val Ala Pro Leu Cys Arg Gly His  
20 25 30  
Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala Trp  
35 40 45  
Asn Arg Asn Leu Ala Asn Gln Arg His Phe Phe Cys Pro Ser Ile Phe  
50 55 60  
His Thr Cys Pro Thr Val Leu Leu Phe Xaa  
65 70

(2) INFORMATION FOR SEQ ID NO: 402:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 402:

Ala Arg Thr Ile Leu Val Leu Tyr Leu Ser Leu Gln Arg Leu Glu Asn  
1 5 10 15  
Leu Ala Tyr His  
20

(2) INFORMATION FOR SEQ ID NO: 403:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 403:

Met Pro Leu Pro Ser Val Pro Ile Leu Gly Ile Phe Ser Phe Leu Ile  
1 5 10 15  
Pro Ser Ser Gln Gly Val Ser Tyr Thr Lys Leu Pro Ile Ser Ser Pro  
20 25 30

Gln Tyr Ser Pro Phe Val Asn Asp His Phe Ser Phe Leu Asn Pro Phe  
35 40 45

Pro Val Gln Ile His Thr Gly Phe Ala Arg Val Gly Ser Tyr Met Gln  
50 55 60

Met Pro Leu Val His Leu Cys Leu Leu Gln Thr Ser Leu Met Lys Asn  
65 70 75 80

Ser Gly Val Gln Gln Gly Ser  
85

596

(2) INFORMATION FOR SEQ ID NO: 404:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 404:

Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile Ile  
1 5 10 15

Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala Val  
20 25 30

Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile Thr  
35 40 45

Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val Tyr  
50 55 60

Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg Asp  
65 70 75 80

Gln Thr Lys Ser Ile Val Glu Lys Ile Pro Ser Lys  
85 90

(2) INFORMATION FOR SEQ ID NO: 405:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 405:

Met Ala Cys Ser Cys Leu Met Ile Gln Ser Phe Ser Thr Ser Ala Leu  
1 5 10 15

Val Leu Phe Tyr Gly  
20

(2) INFORMATION FOR SEQ ID NO: 406:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 174 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 406:

Met Glu Glu Gly Gly Asn Leu Gly Gly Leu Ile Lys Met Val His Leu  
1 5 10 15

Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp Val Thr Phe Val  
20 25 30

Ser Gly Phe Pro Ala Phe Pro Lys Pro Ser Pro Thr Tyr Leu Arg Thr  
35 40 45

5 Ser Ala Glu Gln Thr Leu Pro Leu Leu Pro His Leu His Gly Leu  
50 55 60

Cys Leu His Gln Pro Leu His Leu Gly Phe Thr Ala Cys Leu Gly Ser  
65 70 75 80

10 Ala His Ile Leu Gly Gln Pro Ala Leu Pro Ala Val Pro Glu Pro  
85 90 95

15 Tyr Ala Gly His Cys Gln Arg Pro Leu Ala Gly Thr Pro His His Ser  
100 105 110

Cys His Val Gly Pro Ala Asn Arg Gly Arg Arg Ser Glu Ala Trp Val  
115 120 125

20 Gly Arg Tyr Gln Ala Ala Asn Arg Phe Pro Ile Leu Asn Ala Xaa Cys  
130 135 140

Glu Arg Arg Thr Pro Ser Thr Val Leu Ser Ala Arg Ile Ser Ser Ala  
145 150 155 160

25 Thr Met Gly Cys Pro Leu Phe Ala Ile Trp Ala Ala Ser Xaa  
165 170

30 (2) INFORMATION FOR SEQ ID NO: 407:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 407:

Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu  
1 5 10 15

Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His  
20 25 30

45 Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys  
35 40 45

50 Thr Glu Gly Ile Pro Cys Gln Arg Tyr Gln Asn Gly Leu His Ile Xaa  
55 60

55 (2) INFORMATION FOR SEQ ID NO: 408:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 280 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 408:

Met Glu Ala Val Val Asn Leu Tyr Gln Glu Val Met Lys His Ala Asp  
1 5 10 15

Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser Pro Leu Leu Met Thr  
20 25 30

10 Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu Ser Leu Gly Pro Arg  
35 40 45

Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg Gly Phe Met Ile Val  
50 55 60

15 Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr Ile Val Tyr Glu Phe  
65 70 75 80

Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp Arg Cys Asp Pro Val  
85 90 95

20 Asp Tyr Ser Asn Ser Pro Glu Ala Leu Arg Met Val Arg Val Ala Trp  
100 105 110

25 Leu Phe Leu Phe Ser Lys Phe Ile Glu Leu Met Asp Thr Val Ile Phe  
115 120 125

Ile Leu Arg Lys Lys Asp Gly Gln Val Thr Phe Leu His Val Phe His  
130 135 140

30 His Ser Val Leu Pro Trp Ser Trp Trp Trp Gly Val Lys Ile Ala Pro  
145 150 155 160

35 Gly Gly Met Gly Ser Phe His Ala Met Ile Asn Ser Ser Val His Val  
165 170 175

Ile Met Tyr Leu Tyr Tyr Gly Leu Ser Ala Phe Gly Pro Val Ala Gln  
180 185 190

40 Pro Tyr Leu Trp Trp Lys Lys His Met Thr Ala Ile Gln Leu Ile Gln  
195 200 205

Phe Val Leu Val Ser Leu His Ile Ser Gln Tyr Tyr Phe Met Ser Ser  
210 215 220

45 Cys Asn Tyr Gln Tyr Pro Val Ile Ile His Leu Ile Trp Met Tyr Gly  
225 230 235 240

50 Thr Ile Phe Phe Met Leu Phe Ser Asn Phe Trp Tyr His Ser Tyr Thr  
245 250 255

Lys Gly Lys Arg Leu Pro Arg Ala Leu Gln Gln Asn Gly Ala Pro Gly  
260 265 270

55 Ile Ala Lys Val Lys Ala Asn Xaa  
275 280

60 (2) INFORMATION FOR SEQ ID NO: 409:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 284 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 409:

Met Xaa Leu Trp Pro Gln Thr Cys Ser Gly Lys Phe Asp Gly Thr Leu  
1 5 10 15  
Ala Phe Ser Ile His Xaa Leu Ala Val Ile Leu Gly Asp Gln Leu Thr  
20 25 30  
Ala Ala Asp Leu Val Pro Ile Phe Asn Gly Phe Leu Lys Asp Leu Asp  
35 40 45  
Glu Val Arg Ile Gly Val Leu Lys His Leu His Asp Phe Leu Lys Leu  
50 55 60  
Leu His Ile Asp Lys Arg Arg Glu Tyr Leu Tyr Gln Leu Gln Glu Phe  
65 70 75 80  
Leu Val Thr Asp Asn Ser Arg Asn Trp Arg Phe Arg Ala Glu Leu Ala  
85 90 95  
Glu Gln Leu Ile Leu Leu Leu Glu Leu Tyr Ser Pro Arg Asp Val Tyr  
100 105 110  
Asp Tyr Leu Arg Pro Ile Ala Leu Asn Leu Cys Ala Asp Lys Val Ser  
115 120 125  
Ser Val Arg Trp Ile Ser Tyr Lys Leu Val Ser Glu Met Val Lys Lys  
130 135 140  
Leu His Ala Ala Thr Pro Pro Thr Phe Gly Val Asp Leu Ile Asn Glu  
145 150 155 160  
Leu Val Glu Asn Phe Gly Arg Cys Pro Lys Trp Ser Gly Arg Gln Ala  
165 170 175  
Phe Val Phe Val Cys Gln Thr Val Ile Glu Asp Asp Cys Leu Pro Met  
180 185 190  
Asp Gln Phe Ala Val His Leu Met Pro His Leu Leu Thr Leu Ala Asn  
195 200 205  
Asp Arg Val Pro Asn Val Arg Val Leu Leu Ala Lys Thr Leu Arg Gln  
210 215 220  
Thr Leu Leu Glu Lys Asp Tyr Phe Leu Ala Ser Ala Ser Cys His Gln  
225 230 235 240  
Glu Ala Val Glu Gln Thr Thr Ile Met Ala Leu Gln Met Asp Arg Asp Ser  
245 250 255  
Asp Val Lys Tyr Phe Ala Ser Ile His Pro Ala Ser Thr Lys Ile Ser  
260 265 270  
Glu Asp Ala Met Ser Thr Ala Ser Ser Thr Tyr Xaa  
275 280

## (2) INFORMATION FOR SEQ ID NO: 410:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 187 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 410:

Met Leu Phe Leu Phe Phe Val Ile Ile Phe Leu Phe Val Phe Leu Ile  
1 5 10 15  
Leu Ile Ile Gln Phe Ser Lys Pro Leu Thr Asn Pro His Pro Pro Ala  
20 25 30  
Gly Xaa Ser Asp Arg Arg Arg Tyr Ser Ser Tyr Arg Ser His Asp  
35 40 45  
His Tyr Gln Arg Gln Arg Val Leu Gln Lys Glu Arg Ala Ile Glu Glu  
50 55 60  
Arg Arg Val Val Phe Ile Gly Lys Ile Pro Gly Arg Met Thr Arg Ser  
65 70 75 80  
Glu Leu Lys Gln Arg Phe Ser Val Phe Gly Glu Ile Glu Glu Cys Thr  
85 90 95  
Ile His Phe Arg Val Gln Gly Asp Asn Tyr Gly Phe Val Thr Tyr Arg  
100 105 110  
Tyr Ala Glu Glu Ala Phe Ala Ala Ile Glu Ser Gly His Lys Leu Arg  
115 120 125  
Gln Ala Asp Glu Gln Pro Phe Asp Leu Cys Phe Gly Gly Arg Arg Xaa  
130 135 140  
Xaa Cys Lys Arg Ser Tyr Ser Asp Leu Asp Ser Asn Arg Arg Glu Asp Phe  
145 150 155 160  
Asp Pro Ala Pro Val Lys Ser Lys Phe Asp Ser Leu Asp Phe Asp Thr  
165 170 175  
Leu Leu Lys Gln Ala Gln Lys Asn Leu Arg Arg  
180 185  
(2) INFORMATION FOR SEQ ID NO: 411:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

Met Lys Leu Pro Glu Lys Phe Arg Arg Ala His Gln Gly Asn Leu Glu  
1 5 10 15

Ser Gln Leu Thr Ser Glu Ser Tyr Tyr Lys Glu Thr Leu Ser Val Pro  
 20 25 30  
 5 Thr Val Glu His Ile Ile Gln Glu Lys Asp Ile Phe Ser Glu Gln  
 35 40 45  
 His Leu Lys Ala Leu Lys Cys Leu Ser Leu Val Pro Ser Val Met Gly  
 50 55 60  
 10 Gln Leu Lys Phe Asn Thr Ser Glu His Ala Asp Met Tyr Arg  
 65 70 75 80  
 Ser Asp Leu Pro Asn Pro Asp Thr Leu Ser Ala Glu Leu His Cys Trp  
 85 90 95  
 15 Arg Ile Lys Trp Lys His Arg Gly Lys Asp Ile Glu Leu Pro Ser Thr  
 100 105 110  
 Ile Tyr Glu Ala Leu His Leu Pro Asp Ile Lys Phe Phe Pro Asn Val  
 115 120 125  
 20 Tyr Ala Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu  
 130 135 140  
 Asn Glu Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg  
 145 150 155 160  
 Asn Thr Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile  
 165 170 175  
 30 Asn Phe Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile  
 180 185 190  
 Lys Leu Tyr Thr Xaa Xaa Ser Xaa Leu Xaa Thr Xaa Xaa Ser Xaa Xaa  
 195 200 205  
 Val Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa  
 210 215 220  
 40 Asp Xaa Xaa Xaa Arg Glu Lys Ala Val Arg Cys Met Xaa  
 225 230 235  
 45 (2) INFORMATION FOR SEQ ID NO: 412:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 412:  
 Met Lys Pro Met Ala Val Val Ala Ser Thr Val Leu Gly Leu Val Gln  
 1 5 10 15  
 55 Asn Met Arg Ala Phe Gly Gly Ile Leu Val Val Val Tyr Tyr Val Phe  
 20 25 30  
 Ala Ile Ile Gly Ile Asn Leu Phe Arg Gly Val Ile Val Ala Leu Pro  
 35 40 45

Gly Asn Ser Ser Leu Ala Pro Ala Asn Gly Ser Ala Pro Cys Gly Ser  
 50 55 60  
 5 Phe Glu Gln Leu Glu Tyr Trp Ala Asn Asn Phe Asp Asp Phe Ala Ala  
 65 70 75 80  
 Ala Leu Val Thr Leu Trp Asn Leu Met Val Val Asn Asn Trp Gln Val  
 85 90 95  
 10 Phe Leu Asp Ala Tyr Arg Arg Tyr Ser Gly Pro Trp Ser Lys Ile Tyr  
 100 105 110  
 Phe Val Leu Trp Trp Leu Val Ser Ser Val Ile Trp Val Asn Leu Phe  
 115 120 125  
 15 Leu Ala Leu Ile Leu Glu Asn Phe Leu His Lys Trp Asp Pro Arg Ser  
 130 135 140  
 20 His Leu Gln Pro Leu Ala Gly Thr Pro Glu Ala Thr Tyr Gln Met Thr  
 145 150 155 160  
 Val Glu Leu Leu Phe Arg Asp Ile Leu Glu Glu Pro Gly Glu Asp Glu  
 165 170 175  
 25 Leu Thr Glu Arg Leu Ser Gln His Pro His Leu Trp Leu Cys Arg Xaa  
 180 185 190  
 30  
 35 (2) INFORMATION FOR SEQ ID NO: 413:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 413:  
 Asn Val Val Val Val Ala Phe Gly Leu Ile Leu Ile Ile Glu Ser Leu  
 1 5 10 15  
 45 Gly Glu Gln Cys Pro  
 20  
 50 (2) INFORMATION FOR SEQ ID NO: 414:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 51 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 414:  
 Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys  
 1 5 10 15  
 55 60

Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala  
20 25 30  
Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Leu Leu Lys Met Arg Arg  
35 40 45  
Ala Pro Xaa  
50

(2) INFORMATION FOR SEQ ID NO: 415:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 415:

Met Leu Ile Ile Ser Leu Arg Pro Gln Phe Pro Ser Leu Ile Val Gln  
1 5 10 15  
Leu Glu Cys Ser Val Leu Phe Leu Pro Ile Ser Leu Asn Leu Leu Leu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 416:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 163 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 416:

Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu  
1 5 10 15  
Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu  
20 25 30  
Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp  
35 40 45  
Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr  
50 55 60  
Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu  
65 70 75 80  
Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu  
85 90 95  
Glu Ile Arg Gly Gly Gly Leu Leu Lys Tyr Cys Asn Leu Leu Val Arg  
100 105 110  
Gly Phe Arg Pro Ala Ser Asp Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met

115 120 125  
Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Gln Arg  
130 135 140  
Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Ile Gly Arg Pro  
145 150 155 160  
Gln Val Xaa

(2) INFORMATION FOR SEQ ID NO: 417:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 174 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 417:

Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe  
1 5 10 15  
Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly  
20 25 30  
Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Phe Tyr Ser Ser  
35 40 45  
Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr  
50 55 60  
Gly Ala Ser Tyr His Ser Met Ser Ser Met Ala Arg Ala Ala Phe Ser  
65 70 75 80  
Glu Asp Gly Ala Leu Met Asp Gly Gly Met Asp Leu Asn Met Glu Gln  
85 90 95  
Gly Met Ala Glu His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val  
100 105 110  
Gln Val Leu Ser Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu  
115 120 125  
Ala Pro Gly Arg Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro  
130 135 140  
Trp Phe Thr Ala Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys  
145 150 155 160  
Arg Gln Arg Arg Gln Glu Arg Arg Gln Met Lys Arg Leu Xaa  
165 170

(2) INFORMATION FOR SEQ ID NO: 418:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 50 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 418:

5 Met Glu Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Val Ala Met Met 15

Arg Pro Phe Tyr Leu Leu Pro Val Leu Cys Thr Gln Ala Leu Arg 30

10 Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Cys Cys Leu 45

Ala Xaa 50

(2) INFORMATION FOR SEQ ID NO: 419:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 120 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 419:

25 Met Leu Gly Lys Gly Gly Arg Ala Gly Leu Leu Arg Tyr Arg Leu 15

30 Leu Tyr Phe Thr Leu Val Val Gly Glu Gly Glu Pro Gly Glu Asn Lys 30

Val Thr Ile Pro Phe Phe Glu Thr Gly Lys Lys Ile Ile Phe Cys Ser 45

35 Val Lys Met Val Glu Asn Ser Asn Val Pro Ser His Lys Gly Pro Val 60

Pro Leu Arg Ser Glu Gln Trp Glu Leu Lys Ile Ser Glu Thr Leu Gly 80

Glu Gly Lys Ile Gly Phe Leu Leu Ile Gly Arg Cys Ser Ser Gly Xaa 95

45 Gly Gly Leu Cys Phe Cys Trp Asp Val Leu Cys Cys Met Tyr Ala Tyr 110

Met Asp Arg Ser Leu Ser Leu 120

(2) INFORMATION FOR SEQ ID NO: 420:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 159 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 420:

60

Met Thr His Leu Leu Thr Ala Thr Val Thr Pro Ser Glu Gln Asn 15

Ser Ser Arg Glu Pro Gly Trp Glu Thr Ala Met Ala Lys Asp Ile Leu 30

Gly Glu Ala Gly Leu His Phe Asp Glu Leu Asn Lys Leu Arg Val Leu 45

10 Asp Pro Glu Val Thr Gln Gln Thr Ile Glu Leu Lys Glu Cys Lys 60

Asp Phe Val Asp Lys Ile Gly Gln Phe Gln Lys Ile Val Gly Gly Leu 80

15 Ile Glu Leu Val Asp Gln Leu Ala Lys Glu Ala Glu Asn Glu Lys Met 95

Lys Ala Ile Gly Ala Arg Asn Leu Lys Ser Ile Ala Lys Gln Arg 110

Glu Ala Gln Gln Gln Leu Gln Ala Leu Ile Ala Glu Lys Lys Met 125

25 Gln Leu Glu Arg Tyr Arg Val Glu Tyr Glu Ala Leu Cys Lys Val Glu 140

Ala Glu Gln Asn Glu Phe Ile Asp Gln Phe Ile Phe Gln Lys Xaa 155

(2) INFORMATION FOR SEQ ID NO: 421:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 154 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 421:

Met Asn Val Gly Val Ala His Ser Glu Val Asn Pro Asn Thr Arg Val 15

Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu 30

Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala 45

50 Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu 60

His Ala Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala 80

Arg Leu Leu Thr His Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr 95

Ser Ser Arg Lys Phe Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu 110

Ala Ser Phe Tyr Thr Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr  
115 120 125

5 Ala Ser Leu Leu Ser Val Leu Ile Pro Lys Met Pro Gln Leu His Gly  
130 135 140

Val Arg Ile Phe Gly Ile Asn Lys Tyr Xaa  
145 150

10 (2) INFORMATION FOR SEQ ID NO: 422:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 204 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 422:

20 Met Val Cys Gly Gly Phe Ala Cys Ser Lys Asn Cys Leu Cys Ala Leu  
1 5 10 15

Asn Leu Leu Tyr Thr Leu Val Ser Leu Leu Leu Ile Gly Ile Ala Ala  
20 25 30

Trp Gly Ile Gly Phe Gly Leu Ile Ser Ser Leu Arg Val Val Gly Val  
35 40 45

30 Val Ile Ala Val Ile Phe Leu Phe Leu Ile Ala Leu Val Gly Leu  
50 55 60

Ile Gly Ala Val Lys His His Gln Val Leu Leu Phe Phe Tyr Met Ile  
65 70 75 80

35 Ile Leu Leu Leu Val Phe Ile Val Gln Phe Ser Val Ser Cys Ala Cys  
85 90 95

40 Leu Ala Leu Asn Gln Gln Gln Gly Gln Leu Leu Glu Val Gly Trp  
100 105 110

Asn Asn Thr Ala Ser Ala Arg Asn Asp Ile Gln Arg Asn Leu Asn Cys  
115 120 125

45 Cys Gly Phe Arg Ser Val Asn Pro Asn Asp Thr Cys Leu Ala Ser Cys  
130 135 140

Val Lys Ser Asp His Ser Cys Ser Pro Cys Ala Pro Ile Ile Gly Glu  
145 150 155 160

50 Tyr Ala Gly Glu Val Leu Arg Phe Val Gly Gly Ile Gly Leu Phe Phe  
165 170 175

Ser Phe Thr Glu Ile Leu Gly Val Trp Leu Thr Tyr Arg Tyr Arg Asn  
180 185 190

55 Gln Lys Asp Pro Arg Ala Asn Pro Ser Ala Phe Leu  
195 200

60

(2) INFORMATION FOR SEQ ID NO: 423:

5 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 67 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 423:

10 Met Leu Gln Ser Ile Ile Lys Asn Ile Trp Ile Pro Met Lys Pro Tyr  
1 5 10 15

Tyr Thr Lys Val Tyr Gln Glu Ile Trp Ile Gly Met Gly Leu Met Gly  
20 25 30

15 Phe Ile Val Tyr Lys Ile Arg Ala Ala Asp Lys Arg Ser Lys Ala Leu  
35 40 45

20 Lys Ala Ser Ala Pro Ala Pro Gly His His Asn Gln Ile Tyr Leu Glu  
50 55 60

Tyr Met Xaa  
65

25 (2) INFORMATION FOR SEQ ID NO: 424:

30 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 424:

35 Met Leu Gly Val Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val  
1 5 10 15

40 Ala Val Asn Asn Pro Lys Lys Gln Glu  
20 25

45 (2) INFORMATION FOR SEQ ID NO: 425:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 299 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 425:

50 Met Ala Ala Xaa Glu Pro Ala Val Leu Ala Leu Pro Asn Ser Gly Ala  
1 5 10 15

Gly Gly Ala Gly Ala Pro Ser Gly Thr Val Pro Val Leu Phe Cys Phe  
20 25 30

55 Ser Val Phe Ala Arg Pro Ser Ser Val Pro His Gly Ala Gly Tyr Glu  
35 40 45

60 Leu Leu Ile Gln Lys Phe Leu Ser Leu Tyr Gly Asp Gln Ile Asp Met



WO 98/39448

PCT/US98/04493

WO 98/39448

610

609

50 55 60

His Arg Lys Phe Val Val Gln Leu Phe Ala Glu Glu Thr Gly Gln Tyr  
65 70 75 80

5

Val Asp Leu Pro Lys Gly Phe Ala Val Ser Glu Arg Cys Lys Val Arg  
85 90 95

Leu Val Pro Leu Gln Ile Gln Leu Thr Thr Leu Gly Asn Leu Thr Pro  
100 105 110

10

Ser Ser Thr Val Phe Phe Cys Asp Met Gln Glu Arg Phe Arg Pro  
115 120 125

Ala Ile Lys Tyr Phe Gly Asp Ile Ile Ser Val Gly Gln Arg Leu  
130 135 140

15

Gln Gly Ala Arg Ile Leu Gly Ile Pro Val Ile Val Thr Glu Gln Tyr  
145 150 155 160

20

Pro Lys Gly Leu Gly Ser Thr Val Gln Glu Ile Asp Leu Thr Gly Val  
165 170 175

Lys Leu Val Leu Pro Lys Thr Lys Phe Ser Met Val Leu Pro Glu Val  
180 185 190

25

Glu Ala Ala Leu Ala Glu Ile Pro Gly Val Arg Ser Val Val Leu Phe  
195 200 205

Gly Val Glu Thr His Val Cys Ile Gln Gln Thr Ala Leu Glu Leu Val  
210 215 220

30

Gly Arg Gly Val Glu Val His Ile Val Ala Asp Ala Thr Ser Ser Arg  
225 230 235 240

35

Ser Met Met Asp Arg Met Phe Ala Leu Glu Arg Leu Ala Xaa Gly  
245 250 255

Ile Ile Val Thr Thr Ser Glu Ala Val Leu Leu Gln Leu Val Ala Asp  
260 265 270

40

Lys Asp His Pro Lys Phe Lys Glu Ile Gln Asn Leu Ile Lys Ala Ser  
275 280 285

Ala Pro Glu Ser Gly Leu Leu Ser Lys Val Xaa  
290 295

45

50 (2) INFORMATION FOR SEQ ID NO: 426:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 426:

Met Arg Asp Leu Gly Thr Leu Leu Ser Pro Val Cys Ser  
1 5 10

60

(2) INFORMATION FOR SEQ ID NO: 427:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 427:

Met Phe Gly Cys Leu Val Ala Gly Arg Leu Val Gln Thr Ala Ala Gln  
1 5 10 15

10

Gln Val Ala Glu Asp Lys Phe Val Phe Asp Leu Pro Asp Tyr Glu Ser  
20 25 30

15

Ile Asn His Val Val Phe Met Leu Gly Thr Ile Pro Phe Pro Glu  
35 40 45

Gly Met Gly Gly Ser Val Tyr Phe Ser Tyr Pro Asp Ser Asn Gly Met  
50 55 60

20

Pro Val Trp Gln Leu Leu Gly Phe Val Thr Asn Gly Lys Pro Ser Ala  
65 70 75 80

25

Ile Phe Lys Ile Ser Gly Leu Lys Ser Gly Glu Gly Ser Gln His Pro  
85 90 95

Phe Gly Ala Met Asn Ile Val Arg Thr Pro Ser Val Ala Gln Ile Gly  
100 105 110

30

Ile Ser Val Glu Leu Leu Asp Ser Met Ala Gln Gln Thr Pro Val Gly  
115 120 125

35

Asn Ala Ala Val Ser Ser Val Asp Ser Phe Thr Gln Phe Thr Gln Lys  
130 135 140

Met Leu Asp Asn Phe Tyr Asn Phe Ala Ser Ser Phe Ala Val Ser Gln  
145 150 155 160

40

Ala Gln Met Thr Pro Ser Pro Ser Glu Met Phe Ile Pro Ala Asn Val  
165 170 175

Val Leu Lys Trp Tyr Glu Asn Phe Gln Arg Arg Leu Ala Gln Asn Pro  
180 185 190

45

Xaa Phe Trp Xaa Thr Xaa  
195

50

(2) INFORMATION FOR SEQ ID NO: 428:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 428:

Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser  
1 5 10

60

1 5 10 15  
Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Leu Trp Lys  
20 25 30  
5 Asn Arg Gly Val Gly Arg Ser Val Met Ser Ala Val Gly Xaa  
35 40 45  
10 (2) INFORMATION FOR SEQ ID NO: 429:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 370 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 429:  
Met Lys Lys Val Glu Glu Lys Arg Val Asp Val Asn Ser Ala Val Ala  
1 5 10 15  
Met Gly Glu Val Ile Leu Ala Val Cys His Pro Asp Cys Ile Thr Thr  
20 25 30  
25 Ile Lys His Trp Ile Thr Ile Ile Arg Ala Arg Phe Glu Glu Val Leu  
35 40 45  
Thr Trp Ala Lys Gln His Gln Arg Leu Glu Thr Ala Leu Ser Glu  
50 55 60  
Leu Val Ala Asn Ala Glu Leu Leu Glu Leu Ala Trp Ile Gln  
65 70 75 80  
30 Trp Ala Glu Thr Thr Leu Ile Gln Arg Asp Gln Glu Pro Ile Pro Gln  
85 90 95  
35 Asn Ile Asp Arg Val Lys Ala Leu Ile Ala Glu His Gln Thr Phe Met  
100 105 110  
Glu Glu Met Thr Arg Lys Gln Pro Asp Val Asp Arg Val Thr Lys Thr  
115 120 125  
40 Tyr Lys Arg Lys Asn Ile Glu Pro Thr His Ala Pro Phe Ile Glu Lys  
130 135 140  
45 Ser Arg Ser Gly Arg Lys Ser Leu Ser Gln Pro Thr Pro Pro Pro  
145 150 155 160  
Met Pro Ile Leu Ser Gln Ser Glu Ala Lys Asn Pro Arg Ile Asn Gln  
165 170 175  
50 Leu Ser Ala Arg Trp Gln Gln Val Trp Leu Leu Ala Leu Glu Arg Gln  
180 185 190  
55 Arg Lys Leu Asn Asp Ala Leu Asp Arg Leu Glu Glu Lys Glu Phe  
195 200 205  
Ala Asn Phe Asp Phe Asp Val Trp Arg Lys Lys Tyr Met Arg Trp Met  
210 215 220  
60

Asn His Lys Lys Ser Arg Val Met Asp Phe Phe Arg Arg Ile Asp Lys  
225 230 235 240  
5 Asp Gln Asp Gly Lys Ile Thr Arg Gln Glu Phe Ile Asp Gly Ile Leu  
245 250 255  
Ala Ser Lys Phe Pro Thr Thr Lys Leu Glu Met Thr Ala Val Ala Asp  
260 265 270  
10 Ile Phe Asp Arg Asp Gly Asp Gly Tyr Ile Asp Tyr Tyr Glu Phe Val  
275 280 285  
15 Ala Ala Leu His Pro Asn Lys Asp Ala Tyr Arg Pro Thr Thr Asp Ala  
290 295 300  
Asp Lys Ile Glu Asp Glu Val Thr Arg Gln Val Ala Gln Cys Lys Cys  
305 310 315 320  
20 Ala Lys Arg Phe Gln Val Glu Gln Ile Gly Glu Asn Lys Tyr Arg Phe  
325 330 335  
Phe Leu Gly Asn Gln Phe Gly Asp Ser Gln Gln Leu Leu Val Arg  
340 345 350  
25 Ile Leu Arg Asn Arg Asp Gly Ser Arg Trp Trp Arg Met Asp Gly Leu  
355 360 365  
Gly Xaa  
370  
30 (2) INFORMATION FOR SEQ ID NO: 430:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 430:  
Met Asn Val Lys Thr Phe Ser Xaa Asp His Met His Phe Leu Cys Cys  
1 5 10 15  
45 Leu Tyr Leu Arg Tyr Val Thr Phe Val Tyr Leu Asn Leu Phe  
20 25 30  
50 (2) INFORMATION FOR SEQ ID NO: 431:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 431:  
Met Glu Pro His Leu Arg Cys Arg Val Thr Arg Val Arg Gly Ser Leu  
1 5 10 15  
60 Gly Asn Thr Gly Arg Trp Leu Leu

613

614

20

5 (2) INFORMATION FOR SEQ ID NO: 412:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 412:

Met His Tyr Leu Val Leu Gly Gly Leu Gly Val Phe Leu Phe Ser  
1 5 10 1515 Cys Phe Val Phe Leu Phe Phe Xaa Phe Ser Phe Ala Phe Phe Pro Phe  
20 25 3020 Tyr Leu Glu Gly Met Gly Gly Ser Gly Asn Arg Glu Val Gly Gly Gly  
35 40 45Phe Cys Leu Phe Phe  
50

25

(2) INFORMATION FOR SEQ ID NO: 433:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 176 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 433:

35 Met Val Ser Lys Ala Leu Leu Arg Leu Val Ser Ala Val Asn Arg Arg  
1 5 10 1540 Arg Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val Ala Ser  
20 25 30Val Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu  
35 40 4545 Leu Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu  
50 55 60Lys Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val  
65 70 75 8050 Lys Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Ile Thr Gly Gly  
85 90 95Ala Gly Phe Val Gly Ser His Leu Thr Asp Lys Leu Met Met Asp Gly  
100 105 11055 His Glu Val Thr Val Val Asp Asn Phe Phe Thr Gly Arg Lys Arg Asn  
115 120 12560 Val Glu His Trp Ile Gly His Glu Asn Phe Glu Leu Ile Asn His Asp  
130 135 140Val Trp Ser Pro Ser Thr Arg Leu Thr Arg Tyr Thr Ile Trp His  
145 150 155 1605 Leu Gln Pro Pro Leu Gln Thr Thr Cys Ile Ile Leu Ser Arg His Xaa  
165 170 175

10

(2) INFORMATION FOR SEQ ID NO: 434:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 434:

20 Met Leu Arg Cys Trp Pro Leu Phe Trp Leu Pro Leu Val Ser Pro Phe  
1 5 10 1525 Cys Ser Leu Phe Trp Leu Leu Val Glu Trp Phe Gly Thr Asn Ile Asp  
20 25 3030 Arg Glu Ser Tyr Asp Ala Ile Gly Gly Pro Ser Trp Met Thr Ala Ser  
35 40 45Ser Phe Cys Leu Ser Asn Ser Asn Ile Trp Ser Leu Glu Ile Ser Ser  
50 55 6035 Gly Ser Thr Ser Val Val His Ser Gln Gln Ala Met Asp  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 435:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 435:

45 Met Arg Ser Cys Glu Ile Gln Leu Cys Val Trp Leu Leu Val Ser Ser  
1 5 10 1550 His Val Asp Met Val Leu Gly Gly Ser Pro Ser Thr Leu Tyr Met Met  
20 25 30

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 436:

5 Met Val Val Asn Ser Leu Cys Phe Leu Ser Leu Leu Val Ile Leu  
1 5 10 15  
Glu Leu Ser Thr Asp Ser Ser Ala Arg Leu Leu Tyr His Glu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 437:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 437:

20 Met Asp Lys Gln Lys His Leu Glu Val Arg Arg Ser Val Phe Lys Ile  
1 5 10 15  
Gln Gly Lys Ile Ala Phe Ser Leu Met Phe Val Leu Lys Asp Leu Ser  
20 25 30  
Pro Thr Ile Phe Ser His Ser Ile Leu Leu Leu Leu Pro His His Val  
35 40 45  
Leu Pro Cys Thr Pro Gln Met Val Arg Gly Val Thr Gln Val Leu Arg  
50 55 60  
Glu Phe Gly Asp Gln  
65

(2) INFORMATION FOR SEQ ID NO: 438:

40 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 438:

45 Met Pro Leu Cys Phe Phe Ser Phe Leu Cys Trp Val Leu Val Phe  
1 5 10 15  
Lys Leu Ile  
50

(2) INFORMATION FOR SEQ ID NO: 439:

55 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 439:

Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg  
1 5 10 15

5 Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg  
20 25 30

Arg Val Met Val Asn Leu Asn Ile Leu Phe Xaa  
35 40

(2) INFORMATION FOR SEQ ID NO: 440:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 440:

20 Met Leu Leu Phe Pro Ser Leu Leu Phe Ala Ala Thr Tyr Asn Val Ala  
1 5 10 15  
Asn Pro Ser Arg Leu Ile Leu Tyr Met Ile Ser Ala Gly Ala Asp Ser  
20 25 30  
Gln

(2) INFORMATION FOR SEQ ID NO: 441:

30 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 53 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 441:

40 Met Trp Gln Val Arg Gly Leu Pro Pro Val Pro Leu Leu Thr Met  
1 5 10 15  
Ser Pro Pro Cys Leu Ser Ser Pro Phe Pro Phe Ile Ser Val Pro  
20 25 30  
Leu Phe Glu Ala Val Pro Ile Ser Val Ser Asp Gln Pro Ser Pro Xaa  
35 40 45

Leu Thr Thr Leu Leu  
50

(2) INFORMATION FOR SEQ ID NO: 442:

55 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 442:

60

Met Ile Thr Ser Val Leu Val Phe Leu Ile Phe Phe Pro Tyr Leu  
1 5 10 15  
5 Ser Leu Val Thr Leu Leu Gln Ala Arg Asn Leu Trp Val Ile His Arg  
20 25 30  
Ala Ala Leu Cys Glu Ser Gly Leu Phe His Trp Arg Lys Gly Ile Glu  
35 40 45  
10 Asn Gln Leu Glu Pro Met Tyr Phe Leu Pro His Gly Thr Leu Phe Leu  
50 55 60

## (2) INFORMATION FOR SEQ ID NO: 443:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 443:

Met Leu Tyr Ser Cys Glu Pro Tyr Leu Ile Ile Leu Asn Ile Tyr Ser  
1 5 10 15  
30 Gln Lys Ala Phe Tyr Phe Tyr Phe Glu Gly Ser Phe Ser Val Cys  
20 25 30  
Thr Leu

## (2) INFORMATION FOR SEQ ID NO: 444:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 89 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 444:

Met Arg Gln Arg Gln Ala Ala Cys Gln Pro Pro Ser Arg Asn Gly  
1 5 10 15  
Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Phe Phe Leu Val  
20 25 30  
Lys Leu Leu Phe Ile Pro Trp Leu Ala Ser Leu Leu Ser Ser Pro Leu  
35 40 45  
50 Asn Leu Leu Leu Val Ser Ile Ser Trp Asp Leu Gly Leu Lys Leu  
50 55 60  
Asn Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr  
65 70 75 80 85

Lys Lys Phe Asn Lys Lys Lys Lys Lys  
85

## (2) INFORMATION FOR SEQ ID NO: 445:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 350 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 445:

Met Asp Phe Ile Thr Ser Thr Ala Ile Leu Pro Leu Leu Phe Gly Cys  
1 5 10 15  
Leu Gly Val Phe Gly Leu Phe Arg Leu Leu Gln Trp Val Arg Gly Lys  
20 25 30  
Ala Tyr Leu Arg Asn Ala Val Val Ile Thr Gly Ala Thr Ser Gly  
35 40 45  
Leu Gly Lys Glu Cys Ala Lys Val Phe Tyr Ala Ala Gly Ala Lys Leu  
50 55 60  
Val Leu Cys Gly Arg Asn Gly Gly Ala Leu Glu Glu Leu Ile Arg Glu  
65 70 75 80  
Leu Thr Ala Ser His Ala Thr Lys Val Gln Thr His Lys Pro Tyr Leu  
85 90 95  
Val Thr Phe Asp Leu Thr Asp Ser Gly Ala Ile Val Ala Ala Ala  
100 105 110  
Glu Ile Leu Gln Cys Phe Gly Tyr Val Asp Ile Leu Val Asn Asn Ala  
115 120 125  
Gly Ile Ser Tyr Arg Gly Thr Ile Met Asp Thr Thr Val Asp Val Asp  
130 135 140  
Lys Arg Val Met Glu Thr Asn Tyr Phe Gly Pro Val Ala Leu Thr Lys  
145 150 155 160  
Ala Leu Leu Pro Ser Met Ile Lys Arg Arg Gln Gly His Ile Val Ala  
165 170 175  
Ile Ser Ser Ile Gln Gly Lys Met Ser Ile Pro Phe Arg Ser Ala Tyr  
180 185 190  
Ala Ala Ser Lys His Ala Thr Gln Ala Phe Phe Asp Cys Leu Arg Ala  
195 200 205  
Glu Met Glu Gln Tyr Glu Ile Glu Val Thr Val Ile Ser Pro Gly Tyr  
210 215 220  
Ile His Thr Asn Leu Ser Val Asn Ala Ile Thr Ala Asp Gly Ser Arg  
225 230 235 240  
Tyr Gly Val Met Asp Thr Thr Ala Gln Gly Arg Ser Pro Val Glu  
245 250 255

Val Ala Gln Asp Val Leu Ala Val Gly Lys Lys Lys Asp Val  
260 265 270

Ile Leu Ala Asp Leu Leu Pro Ser Leu Ala Val Tyr Leu Arg Thr Leu  
275 280 285

Ala Pro Gly Leu Phe Phe Ser Leu Met Pro Pro Gly Pro Glu Lys Ser  
290 295 300

Gly Asn Pro Arg Thr Pro Ser Thr Leu Thr Ser Gln Gly Gln Gly Arg  
305 310 315 320

Glu Ala Ala Leu Leu Gly Leu Leu Thr Leu Gln Gly Thr Val Ala Phe  
325 330 335

Val Glu Thr Leu Met Glu Ile Cys Leu Thr Ser Gly Lys Asp  
340 345 350

(2) INFORMATION FOR SEQ ID NO: 446:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 49 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 446:

Met Val Phe Leu Pro Arg Gly Val Val Ser Gly Gly Ala Ala Cys  
1 5 10 15

Leu Trp Leu Thr Phe Ile Leu Glu Thr Glu Val Tyr Leu Asp Leu Ala  
20 25 30

Thr Glu Ala Arg Ala His Ser Arg Met Gly Leu Gly Leu Trp Pro Pro  
35 40 45

Asn  
40

(2) INFORMATION FOR SEQ ID NO: 447:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 278 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 447:

Met Ala Ser Ala Glu Leu Asp Tyr Thr Ile Glu Ile Pro Asp Gln Pro  
1 5 10 15

Cys Trp Ser Gln Lys Asn Ser Pro Ser Pro Gly Gly Lys Glu Ala Glu  
20 25 30

Thr Arg Gln Pro Val Val Ile Leu Leu Gly Trp Gly Gly Cys Lys Asp  
35 40 45 50

Lys Asn Leu Ala Lys Tyr Ser Ala Ile Tyr His Lys Arg Gly Cys Ile  
50 55 60

Val Ile Arg Tyr Thr Ala Pro Trp His Met Val Phe Phe Ser Glu Ser  
65 70 75 80

Leu Gly Ile Pro Ser Leu Arg Val Leu Ala Gln Lys Leu Leu Glu Leu  
85 90 95

Leu Phe Asp Tyr Glu Ile Glu Lys Glu Pro Leu Leu Phe His Val Phe  
100 105 110

Ser Asn Gly Gly Val Met Leu Tyr Arg Tyr Val Leu Glu Leu Gln  
115 120 125

Thr Arg Arg Phe Cys Arg Leu Arg Val Val Gly Thr Ile Phe Asp Ser  
130 135 140

Ala Pro Gly Asp Ser Asn Leu Val Gly Ala Leu Arg Ala Ala Ala  
145 150 155 160

Ile Leu Glu Arg Arg Ala Ala Met Leu Arg Leu Leu Leu Val Ala  
165 170 175

Phe Ala Leu Val Val Val Leu Phe His Val Leu Leu Ala Pro Ile Thr  
180 185 190

Ala Xaa Phe His Thr His Phe Tyr Asp Arg Leu Gln Asp Ala Gly Ser  
195 200 205

Arg Trp Pro Glu Leu Tyr Leu Tyr Ser Arg Ala Asp Glu Val Val Leu  
210 215 220

Ala Arg Asp Ile Glu Arg Met Val Glu Ala Arg Leu Ala Arg Arg Val  
225 230 235 240

Leu Ala Arg Ser Val Asp Phe Val Ser Ser Ala His Val Ser His Leu  
245 250 255

Arg Asp Tyr Pro Thr Tyr Tyr Thr Ser Leu Cys Val Asp Phe Met Arg  
260 265 270

Asn Cys Val Arg Cys Xaa  
275

(2) INFORMATION FOR SEQ ID NO: 448:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 199 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 448:

Met Ser Phe Ile Phe Asp Trp Ile Tyr Ser Gly Phe Ser Ser Val Leu  
1 5 10 15

Gln Phe Leu Gly Leu Tyr Lys Lys Thr Gly Lys Leu Val Phe Leu Gly  
20 25 30

621

622

Leu Asp Asn Ala Gly Lys Thr Thr Leu Leu His Met Leu Lys Asp Asp  
35 40 45

5 Arg Leu Gly Gln His Val Pro Thr Leu His Pro Thr Ser Glu Glu Leu  
50 55 60

Thr Ile Ala Gly Met Thr Thr Thr Phe Asp Leu Gly Gly His Val  
65 70 75 80

10 Gln Ala Arg Arg Val Trp Lys Asn Tyr Leu Pro Ala Ile Asn Gly Ile  
85 90 95

15 Val Phe Leu Val Asp Cys Ala Asp His Glu Arg Leu Leu Glu Ser Lys  
100 105 110

Glu Glu Leu Asp Ser Leu Met Thr Asp Glu Thr Ile Ala Asn Val Pro  
115 120 125

20 Ile Leu Ile Leu Gly Asn Lys Ile Asp Arg Pro Glu Ala Ile Ser Glu  
130 135 140

Glu Arg Leu Arg Glu Met Phe Gly Leu Tyr Gly Gln Thr Thr Gly Lys  
145 150 155 160

25 Gly Ser Ile Ser Leu Lys Glu Leu Asn Ala Arg Pro Leu Glu Val Phe  
165 170 175

30 Met Cys Ser Val Leu Lys Arg Gln Gly Tyr Gly Glu Gly Phe Arg Trp  
180 185 190

Met Ala Gln Tyr Ile Asp Xaa  
195

35

(2) INFORMATION FOR SEQ ID NO: 449:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 258 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 449:

45 Met Thr Leu Ser Arg Phe Ala Tyr Asn Gly Lys Arg Cys Pro Ser Ser  
1 5 10 15

Tyr Asn Ile Leu Asp Asn Ser Lys Ile Ile Ser Glu Glu Cys Arg Lys  
20 25 30

50 Glu Leu Thr Ala Leu Leu His His Tyr Tyr Pro Ile Glu Ile Asp Pro  
35 40 45

55 His Arg Thr Val Lys Glu Lys Leu Pro His Met Val Glu Trp Trp Thr  
50 55 60

Lys Ala His Asn Leu Leu Cys Gln Gln Lys Ile Gln Lys Phe Gln Ile  
65 70 75 80

60 Ala Gln Val Val Arg Glu Ser Asn Ala Met Leu Arg Glu Gly Tyr Lys  
75 80

622

85 90 95

Thr Phe Phe Asn Thr Leu Tyr His Asn Asn Ile Pro Leu Phe Ile Phe  
100 105 110

5 Ser Ala Gly Ile Gly Asp Ile Leu Glu Glu Ile Ile Arg Gln Met Lys  
115 120 125

10 Val Phe His Pro Asn Ile His Ile Val Ser Asn Tyr Met Asp Phe Asn  
130 135 140

Glu Asp Gly Phe Leu Gln Gly Phe Lys Gly Gln Leu Ile His Thr Tyr  
145 150 155 160

15 Asn Lys Asn Ser Ser Val Cys Glu Asn Xaa Gly Tyr Phe Gln Gln Leu  
165 170 175

Glu Gly Lys Thr Asn Val Ile Leu Leu Gly Asp Ser Ile Gly Asp Leu  
180 185 190

20 Thr Met Ala Asp Gly Val Pro Gly Val Gln Asn Ile Leu Lys Ile Gly  
195 200 205

Phe Leu Asn Asp Lys Val Glu Glu Arg Arg Xaa Arg Tyr Met Asp Ser  
210 215 220

Tyr Asp Ile Val Leu Glu Lys Asp Glu Thr Leu Asp Val Val Asn Gly  
225 230 235 240

30 Leu Leu Gln His Ile Leu Cys Gln Gly Val Gln Leu Glu Met Gln Gly  
245 250 255

Pro Xaa

35

(2) INFORMATION FOR SEQ ID NO: 450:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 450:

45 Met Ser His Val Leu Leu Cys Pro Ser Leu Ser Cys Ser Asn Leu Leu  
1 5 10 15

Pro Pro Ser His Ser Leu Gly Thr Met Gly Ser Leu Ser Pro His Leu  
20 25 30

Cys Gly His Thr Met Cys Pro Val Asn Pro Glu Leu Pro Leu Ser Ser  
35 40 45

55 Arg Leu Thr Thr Asp Gln Pro Gln Pro Asp Ala Cys Ser Pro Thr Leu  
50 55 60

Leu Thr Leu Pro Leu Pro Ser Ser Phe Leu Pro His Ser Lys Pro Thr  
65 70 75 80

Phe Xaa His Pro Cys Ser Pro  
85

5 (2) INFORMATION FOR SEQ ID NO: 451:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 315 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 451:

15 Met Phe Ser Ile Asn Pro Leu Glu Asn Leu Lys Val Tyr Ile Ser Ser  
1 5 10 15

Arg Pro Pro Leu Val Val Phe Met Ile Ser Val Xaa Pro Met Ala Ile  
20 25 30

20 Ala Phe Leu Thr Leu Gly Tyr Phe Phe Lys Ile Lys Glu Ile Lys Ser  
35 40 45

Pro Glu Met Ala Glu Asp Trp Asn Thr Phe Leu Leu Arg Phe Asn Asp  
50 55 60

25 Leu Asp Leu Cys Val Ser Glu Asn Glu Thr Leu Lys His Leu Thr Asn  
65 70 75 80

Asp Thr Thr Thr Pro Glu Ser Thr Met Thr Ser Gly Gln Ala Arg Ala  
85 90 95

Ser Thr Gln Ser Pro Gln Ala Leu Glu Asp Ser Gly Pro Val Asn Ile  
100 105 110

35 Ser Val Ser Ile Thr Leu Thr Leu Asp Pro Leu Lys Pro Phe Gly Gly  
115 120 125

Tyr Ser Arg Asn Val Thr His Leu Tyr Ser Thr Ile Leu Gly His Gln  
130 135 140

Ile Gly Leu Ser Gly Arg Glu Ala His Glu Glu Ile Asn Ile Thr Phe  
145 150 155 160

45 Thr Leu Pro Thr Ala Trp Ser Ser Asp Asp Cys Ala Leu His Gly His  
165 170 175

Cys Glu Gln Val Val Phe Thr Ala Cys Met Thr Leu Thr Ala Ser Pro  
180 185 190

50 Gly Val Phe Pro Val Thr Val Gln Pro Pro His Cys Val Pro Asp Thr  
195 200 205

Tyr Ser Asn Ala Thr Leu Trp Tyr Lys Ile Phe Thr Thr Ala Arg Asp  
210 215 220

55 Ala Asn Thr Lys Tyr Ala Gln Asp Tyr Asn Pro Phe Trp Cys Tyr Lys  
225 230 235 240

60 Gly Ala Ile Gly Lys Val Tyr His Ala Leu Asn Pro Lys Leu Thr Val  
245 250 255

Ile Val Pro Asp Asp Asp Arg Ser Leu Ile Asn Leu His Leu Met His  
260 265 270

5 Thr Ser Tyr Phe Leu Phe Val Met Val Ile Thr Met Phe Cys Tyr Ala  
275 280 285

Val Ile Lys Gly Arg Pro Ser Lys Leu Arg Gln Ser Asn Pro Glu Phe  
290 295 300

10 Cys Pro Glu Lys Val Ala Leu Ala Glu Ala Xaa  
305 310 315

15 (2) INFORMATION FOR SEQ ID NO: 452:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 452:

25 Met Pro Gly Leu Ser Leu Ala Leu Leu Pro Phe Gly Pro Gly Cys Thr  
1 5 10 15

Glu Ala Leu His Ala Gly Cys Phe Pro Ala Phe Ala Ser Ala Thr Arg  
20 25 30

30 Val Asn Gly Glu Ala Ala Leu Ser Pro Gly Leu Cys Asp Pro Ile Ser  
35 40 45

Val Pro Tyr Val  
50

35 (2) INFORMATION FOR SEQ ID NO: 453:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 383 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 453:

45 Met Ala Val Gly Gln Ile Met Thr Phe Gly Ser Pro Val Ile Gly Cys  
1 5 10 15

50 Gly Phe Ile Ser Gly Trp Asn Leu Val Ser Met Cys Val Glu Tyr Val  
20 25 30

Leu Leu Trp Lys Val Tyr Gln Lys Thr Pro Ala Leu Ala Val Lys Ala  
35 40 45

55 Gly Leu Lys Glu Glu Glu Thr Glu Leu Lys Gln Leu Asn Leu His Lys  
50 55 60

Asp Thr Glu Pro Lys Pro Leu Glu Gly Thr His Leu Met Gly Val Lys  
65 70 75 80



Asp Ser Asn Ile His Glu Leu Glu His Glu Gln Glu Pro Thr Cys Ala 95  
 5 Ser Gln Met Ala Glu Pro Phe Arg Thr Phe Arg Asp Gly Trp Val Ser 100  
 Tyr Tyr Asn Gln Pro Val Phe Leu Ala Gly Met Gly Leu Ala Phe Leu 115  
 10 Tyr Met Thr Val Leu Gly Phe Asp Cys Ile Thr Thr Gly Tyr Ala Tyr 130  
 Thr Gln Gly Leu Ser Gly Phe His Pro Gln Tyr Phe Asp Gly Ser Ile 145  
 15 Ser Tyr Asn Trp Asn Asn Gly Asn Cys Ser Phe Tyr Leu Ala Thr Ser 165  
 Lys Met Trp Phe Gly Ser Ala Gly Leu Ile Ser Gly Leu Ala Gln Leu 180  
 20 Ser Cys Leu Ile Leu Cys Val Ile Ser Val Phe Met Pro Gly Ser Pro 195  
 Leu Asp Leu Ser Val Ser Pro Phe Glu Asp Ile Arg Ser Arg Phe Ile 210  
 25 Gln Gly Glu Ser Ile Thr Pro Thr Lys Ile Pro Glu Ile Thr Thr Glu 225  
 Ile Tyr Met Ser Asn Gly Ser Asn Ser Ala Asn Ile Val Pro Glu Thr 245  
 30 Ser Pro Glu Ser Val Pro Ile Ile Ser Val Ser Leu Leu Phe Ala Gly 260  
 Val Ile Ala Ala Arg Ile Gly Leu Trp Ser Phe Asp Leu Thr Val Thr 275  
 35 Gln Leu Leu Gln Glu Asn Val Ile Glu Ser Glu Arg Gly Ile Ile Asn 290  
 Gly Val Gln Asn Ser Met Asn Tyr Leu Leu Asp Leu Leu His Phe Ile 305  
 40 Met Val Ile Leu Ala Pro Asn Pro Glu Ala Phe Gly Leu Leu Val Leu 325  
 Ile Ser Val Ser Phe Val Ala Met Gly His Ile Met Tyr Phe Arg Phe 340  
 45 Ala Gln Asn Thr Leu Gly Asn Lys Leu Phe Ala Cys Gly Pro Asp Ala 355  
 Lys Glu Val Arg Lys Glu Asn Gln Ala Asn Thr Ser Val Val Xaa 370  
 55

60 (2) INFORMATION FOR SEQ ID NO: 454:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 186 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 454:  
 Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe 1  
 5 Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met 20  
 10 Met Trp Ala Ala Gly Ala Val Ala Met Ser Ser Ile Thr Phe Pro 35  
 15 Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly 50  
 20 Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu 65  
 Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu 85  
 25 Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro 100  
 30 Gln His His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu 115  
 Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile 130  
 35 Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Trp Arg Lys His 145  
 Cys Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala 165  
 40 Lys Glu Pro Leu Leu Gln Asp Thr Asn Val 180  
 45 (2) INFORMATION FOR SEQ ID NO: 455:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 163 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 455:  
 Met Leu Gln Thr Ser Asn Tyr Ser Leu Val Leu Ser Leu Gln Phe Leu 1  
 5 Leu Leu Ser Tyr Asp Leu Phe Val Asn Ser Phe Ser Glu Leu Leu Gln 20  
 25 Lys Thr Pro Val Ile Gln Leu Val Leu Phe Ile Ile Gln Asp Ile Ala 30

35 40 45  
Val Leu Phe Asn Ile Ile Ile Phe Leu Met Phe Phe Asn Thr Phe  
50 55 60  
Val Phe Gln Ala Gly Leu Val Asn Leu Leu Phe His Lys Phe Lys Gly  
65 70 75 80  
Thr Ile Ile Leu Thr Ala Val Tyr Phe Ala Leu Ser Ile Ser Leu His  
10 85 90 95  
Val Trp Val Met Asn Leu Arg Trp Lys Asn Ser Asn Ser Phe Ile Trp  
100 105 110  
Thr Asp Gly Leu Gln Met Leu Phe Val Phe Gln Arg Leu Ala Val  
115 120 125  
Leu Tyr Cys Tyr Phe Tyr Lys Arg Thr Ala Val Arg Leu Gly Asp Pro  
130 135 140  
His Phe Tyr Gln Asp Ser Leu Trp Leu Arg Lys Gln Phe Met Gln Val  
145 150 155 160  
Arg Arg Xaa

25  
(2) INFORMATION FOR SEQ ID NO: 456:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 46 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 456:  
Met Arg Ile Gln Val Phe Ile Leu Leu Leu Gly Ala Gly Gly Thr Ser  
1 5 10 15  
Gln Phe Thr Lys Pro Pro Ser Leu Pro Leu Gln Pro Gln Pro Ala Val  
20 25 30  
Glu Ser Ser Pro Thr Glu Thr Ser Glu Gln Ile Arg Glu Lys  
35 40 45

45  
(2) INFORMATION FOR SEQ ID NO: 457:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 105 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 457:  
Met Ser Tyr Leu Ala Phe Leu Tyr Met Thr Phe Asp Phe Cys Cys Leu  
1 5 10 15  
Tyr Phe Ser Thr Val Tyr Ala Pro Ser Phe Lys Tyr Ile Cys Val His  
20 25 30

50  
(2) INFORMATION FOR SEQ ID NO: 457:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 105 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 457:  
Met Ser Tyr Leu Ala Phe Leu Tyr Met Thr Phe Asp Phe Cys Cys Leu  
1 5 10 15  
Tyr Phe Ser Thr Val Tyr Ala Pro Ser Phe Lys Tyr Ile Cys Val His  
20 25 30

Thr Asp Thr His Ile Cys Val Cys Val Cys Ile Tyr Leu Ser Ser Val  
35 40 45  
Val Ser Lys Ser Ser Ala Glu Ala Asp Gly Val Leu Gln Pro Arg Arg  
50 55 60  
His Pro Ala Ser Leu Leu Ile Val Phe Ala Thr Ser Ile Ser Glu Ser  
65 70 75 80  
Ser Leu Leu Ile Phe Ser Phe Gln Lys Thr Glu Ala Lys Leu Ile Val  
85 90 95  
Phe Ala Val Ser Leu Ala Ala Lys Xaa  
100 105

15  
(2) INFORMATION FOR SEQ ID NO: 458:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 458:  
Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser  
1 5 10 15  
Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val  
20 25 30  
Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro  
35 40 45  
Ser Ser Glu Leu Leu Gln Gln Leu Leu Ser Val Gln Phe Val Trp Gln  
50 55 60  
Ala His Thr Val Ala Xaa  
65 70

40  
(2) INFORMATION FOR SEQ ID NO: 459:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 155 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 459:  
Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe  
1 5 10 15  
Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30  
Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45

50  
(2) INFORMATION FOR SEQ ID NO: 459:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 155 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 459:  
Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe  
1 5 10 15  
Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30  
Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45

60  
(2) INFORMATION FOR SEQ ID NO: 459:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 155 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 459:  
Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe  
1 5 10 15  
Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30  
Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45

Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala  
50 55 60

5 Val Gly Phe Leu Glu Leu Leu Ala Gly Leu Leu Val Met Gly Pro  
65 70 75 80  
Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met  
85 90 95

10 Gly Ala Ile Phe Thr Leu Ala Leu Lys Glu Ser Leu Ser Thr Cys  
100 105 110

Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Asn Val Gly  
115 120 125

15 Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys  
130 135 140

20 Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa  
145 150 155

25 (2) INFORMATION FOR SEQ ID NO: 460:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 460:

Met Lys Leu Gly Arg Ala Val Leu Gly Leu Leu Leu Ala Pro Ser  
1 5 10 15

35 Val Val Gln Ala Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly  
20 25 30

40 Val Leu Thr Gly Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu  
35 40 45

Cys Cys Gly Gln Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp  
50 55 60

45 Leu Asp Asp Asn Leu Phe Gly Gln His Leu Ala Lys Lys Ile Ile Leu  
65 70 75 80

Asn Ala Val Phe Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu  
85 90 95

50 Thr Leu Ser Leu His Gly Trp Thr Thr Gly Thr Gly Lys Asn Phe Val Ser  
100 105 110

Lys Ile Ile Ala Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr  
115 120 125

55 Val His Leu Phe Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile  
130 135 140

60 Thr Leu Tyr Lys Asp Gln Leu Leu Trp Ile Arg Gly Asn Val Ser  
145 150 155 160

Ala Cys Ala Arg Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His  
165 170 175

5 Ala Gly Leu Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu  
180 185 190

Val Asp Gly Val Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn  
195 200 205

10 Ala Gly Ala Glu Arg Ile Thr Asp Val Ala Leu Asp Phe Trp Arg Ser  
210 215 220

15 Gly Lys Gln Arg Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu  
225 230 235 240

Ser Val Ser Val Phe Asn Asn Lys Asn Ser Gly Phe Trp His Ser Ser  
245 250 255

20 Leu Ile Asp Arg Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu  
260 265 270

Glu Tyr Lys His Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg  
275 280 285

25 Gly Tyr Glu Ile Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met  
290 295 300

30 Thr Phe Pro Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys  
305 310 315 320

Thr Val Phe Thr Lys Leu Asp Tyr Tyr Asp Asp  
325 330

(2) INFORMATION FOR SEQ ID NO: 461:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 461:

45 Met Leu Lys Cys Ile  
1 5

50 (2) INFORMATION FOR SEQ ID NO: 462:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 462:

Met Ile Leu Thr Leu Leu Ser Val Val Ser Thr Met Ala Ser  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 463:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 285 amino acids  
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 463:

10 Met Lys Leu His Pro Pro Pro Val Thr Gln Asp His Arg  
1 5 10 15  
15 Ser Lys Ser Ser His Ser Asn Trp Met Pro Arg Met Gly Ala Cys Ser  
20 25 30  
Met Ser Arg Thr Ser Ser Ser Gly Pro Pro Ser Leu Cys Lys Ser Thr  
35 40 45  
20 Ser Gly Arg Ser Cys Thr Arg Pro His Cys Trp Pro Ser Leu Pro Ala  
50 55 60  
Trp Val Ser Val Phe Thr Arg Thr Asn Thr Gly Ser Trp Cys Tyr Pro  
65 70 75 80  
25 Ala Trp Gly Gly Ala Phe Ser Arg Pro Trp Met Ser Ala Gln Ser Met  
85 90 95  
30 Cys Cys Ala Gln Arg Ser Val Leu Gln Val Ala Cys Arg Leu Leu Asp  
100 105 110  
Ala Leu Gln Phe Leu His Gln Asn Gln Tyr Val His Gly Asn Val Thr  
115 120 125  
35 Ala Gln Asn Ile Phe Val Asp Pro Gln Asp Gln Val Thr Leu  
130 135 140  
Ala Gly Tyr Gly Phe Ala Phe Arg Tyr Cys Pro Ser Gly Lys His Val  
145 150 155 160  
40 Ala Tyr Val Gln Gly Ser Arg Ser Pro His Gln Gly Asp Leu Gln Phe  
165 170 175  
45 Ile Ser Met Asp Leu His Lys Gly Cys Gly Pro Ser Arg Arg Xaa Asp  
180 185 190  
Leu Gln Ser Leu Gly Tyr Cys Met Leu Lys Trp Leu Tyr Gly Phe Leu  
195 200 205  
50 Pro Trp Thr Asn Cys Leu Pro Xaa Xaa Gln Asp Ile Met Lys Gln Lys  
210 215 220  
Gln Lys Phe Val Asp Lys Pro Gly Pro Phe Val Gly Pro Cys Gly His  
225 230 235 240  
55 Trp Ile Arg Pro Ser Gln Thr Leu Gln Lys Tyr Leu Lys Val Val Met  
245 250 255  
60 Ala Leu Thr Tyr Gln Gln Lys Pro Pro Tyr Ala Met Leu Arg Asn  
260 265 270

Leu Gln Ala Leu Leu Gln Asp Leu Arg Val Ser Pro Tyr  
275 280 285

## (2) INFORMATION FOR SEQ ID NO: 464:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 464:

15 Met Thr Ser Pro Pro His Gln Gly Trp Gln Gln Arg Gly Cys Gly  
1 5 10 15  
Gln Ser Gln Val Pro Leu Ala Leu Ser Arg Val Phe Ser Thr Ser His  
20 25 30  
Tyr Cys Leu Leu Leu Val Ala Asn Gln Ser Ile Phe Phe Pro Cys Leu  
35 40 45  
25 Trp Ala Val Gln Arg Leu Leu Gly Val Arg Cys Thr Cys Pro Leu Ser  
50 55 60  
Trp Gly Lys Arg Ile Ile Ser Gln His Cys Ser Ala Gln Ser Ser Xaa  
65 70 75 80

## (2) INFORMATION FOR SEQ ID NO: 465:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 465:

45 Met His Thr Trp Tyr Asn Asp Arg Arg Gln Asn Cys His Cys Leu Leu  
1 5 10 15  
Phe Phe Leu Ile Tyr Leu Arg Lys Ile Tyr Gln Val Val Pro His Val  
20 25 30  
50 Pro Leu Leu Val Lys Cys Arg Gly Arg Leu Lys Gly Val Asn Ile  
35 40 45

## (2) INFORMATION FOR SEQ ID NO: 466:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 466:

Met Glu Leu Val Leu Val Phe Leu Cys Ser Leu Leu Ala Pro Met Val  
1 5 10 15

5 Leu Ala Ser Ala Ala Glu Lys Glu Lys Glu Met Asp Pro Phe His Tyr  
20 25 30

Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu Val Phe Ala Val Val Leu  
35 40 45

10 Phe Ser Val Gly Ile Leu Leu Ile Leu Ser Arg Cys Lys Cys Ser  
50 55 60

15 Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp Glu Glu Ala Gln Val Glu  
65 70 75 80

Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro Gln Lys Ala Glu Asn Xaa  
85 90 95

20

25 (2) INFORMATION FOR SEQ ID NO: 467:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 399 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 467:

30 Met Ala Ser Gly Ala Asp Ser Lys Gly Asp Asp Leu Ser Thr Ala Ile  
1 5 10 15

Leu Lys Gln Lys Asn Arg Pro Asn Arg Leu Ile Val Asp Glu Ala Ile  
20 25 30

35 Asn Glu Asp Asn Ser Val Val Ser Leu Ser Gln Pro Lys Met Asp Glu  
35 40 45

Leu Gln Leu Phe Arg Gly Asp Thr Val Leu Leu Lys Gly Lys Arg  
50 55 60

45 Arg Glu Ala Val Cys Ile Val Leu Ser Asp Asp Thr Cys Ser Asp Glu  
65 70 75 80

Lys Ile Arg Met Asn Arg Val Val Arg Asn Asn Leu Arg Val Arg Leu  
85 90 95

50 Gly Asp Val Ile Ser Ile Gln Pro Cys Pro Asp Val Lys Tyr Gly Lys  
100 105 110

55 Arg Ile His Val Leu Pro Ile Asp Asp Thr Val Glu Gly Ile Thr Gly  
115 120 125

Asn Leu Phe Glu Val Tyr Leu Lys Pro Tyr Phe Leu Glu Ala Tyr Arg  
130 135 140

60 Pro Ile Arg Lys Gly Asp Ile Phe Leu Val Arg Gly Met Arg Ala

145 150 155 160

Val Glu Phe Lys Val Val Glu Thr Asp Pro Ser Pro Tyr Cys Ile Val  
165 170 175

5 Ala Pro Asp Thr Val Ile His Cys Glu Gly Glu Pro Ile Lys Arg Glu  
180 185 190

10 Asp Glu Glu Ser Leu Asn Glu Val Gly Tyr Asp Asp Ile Gly Gly  
195 200 205

Cys Arg Lys Gln Leu Ala Gln Ile Lys Glu Met Val Glu Leu Pro Leu  
210 215 220

15 Arg His Pro Ala Leu Phe Lys Ala Ile Gly Val Lys Pro Arg Gly  
225 230 235 240

Ile Leu Leu Tyr Gly Pro Gly Thr Gly Lys Thr Leu Ile Ala Arg  
245 250 255

20 Ala Val Ala Asn Glu Thr Gly Ala Phe Phe Leu Ile Asn Gly Pro  
260 265 270

Glu Ile Met Ser Lys Leu Ala Gly Glu Ser Glu Ser Asn Leu Arg Lys  
275 280 285

25 Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro Ala Ile Ile Phe Ile Asp  
290 295 300

30 Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu Lys Thr His Gly Glu Val  
305 310 315 320

Glu Arg Arg Ile Val Ser Gln Leu Leu Thr Leu Met Asp Gly Leu Lys  
325 330 335

35 Gln Arg Ala His Val Ile Val Met Ala Ala Thr Asn Arg Pro Asn Ser  
340 345 350

Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg Phe Asp Arg Glu Val Asp  
355 360 365

40 Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu Glu Ile Leu Gln Ile His  
370 375 380

45 Thr Lys Asn Met Lys Leu Ala Asp Asp Val Asp Leu Glu Gln Xaa  
385 390 395

50 (2) INFORMATION FOR SEQ ID NO: 468:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 468:

55 Leu  
1

60

## (2) INFORMATION FOR SEQ ID NO: 469:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 273 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 469:

10 Met Ala Ala Pro Lys Gly Ser Leu Trp Val Arg Thr Gln Leu Gly Leu  
1 5 10 15  
15 Pro Pro Leu Leu Leu Thr Met Ala Leu Ala Gly Gly Ser Gly Thr  
20 25 30  
Ala Ser Ala Gln Ala Phe Asp Ser Val Leu Gly Asp Thr Ala Ser Cys  
35 40 45  
20 His Arg Ala Cys Gln Leu Thr Tyr Pro Leu His Thr Tyr Pro Lys Gln  
50 55 60  
25 Gln Gln Leu Tyr Ala Cys Gln Arg Gly Cys Arg Leu Phe Ser Ile Cys  
65 70 75 80  
Gln Phe Val Asp Asp Gly Ile Asp Leu Asn Arg Thr Lys Leu Gln Cys  
85 90 95  
30 Gln Ser Ala Cys Thr Gln Ala Tyr Ser Gln Ser Asp Gln Gln Tyr Ala  
100 105 110  
Cys His Leu Gly Cys Gln Asn Gln Leu Pro Phe Ala Gln Leu Arg Gln  
115 120 125  
35 Gln Gln Leu Met Ser Leu Met Pro Lys Met His Leu Leu Phe Pro Leu  
130 135 140  
Thr Leu Val Arg Ser Phe Trp Ser Asp Met Met Asp Ser Ala Gln Ser  
145 150 155 160  
40 Phe Ile Thr Ser Ser Trp Thr Phe Tyr Leu Gln Ala Asp Asp Gly Lys  
165 170 175  
45 Ile Val Ile Phe Xaa Ser Lys Pro Arg Asn Pro Arg Tyr Ala Pro His  
180 185 190  
Leu Gln Pro Gly Ala Leu Pro Asn Leu Xaa Xaa Xaa Ser Leu Ser Lys  
195 200 205  
50 Met Ser Xaa Xaa Ser Xaa Met Arg Asn Ser Gln Ala His Arg Asn Phe  
210 215 220  
Leu Gln Asp Gly Gln Ser Asp Gly Phe Leu Arg Cys Leu Ser Leu Asn  
225 230 235 240  
55 Ser Gly Trp Ile Leu Thr Thr Thr Leu Val Leu Ser Val Met Val Leu  
245 250 255  
60 Leu Trp Ile Cys Cys Ala Thr Cys Cys Tyr Thr Leu Leu Asp Ala Val  
260 265 270

Xaa

## (2) INFORMATION FOR SEQ ID NO: 470:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 470:

15 Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe Ser  
1 5 10 15  
20 Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr Pro  
25 30  
Phe Ile Gly Pro Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile Ala  
35 40 45  
25 Thr Gln Lys Arg Leu Thr Lys Leu Leu Val His Ser Ser Leu Val Gly  
50 55 60  
30 Ser Ile Leu Ser Ala Leu Ser Ala Leu Val Gly Phe Ile Ile Leu Ser  
65 70 75 80  
Val Lys Gln Ala Thr Leu Asn Pro Ala Ser Leu Gln Cys Gln Leu Asp  
85 90 95  
35 Lys Asn Asn Ile Pro Thr Arg Ser Tyr Val Ser Tyr Phe Tyr His Asp  
100 105 110  
Ser Leu Tyr Thr Thr Asp Cys Tyr Thr Ala Lys Ala Ser Leu Ala Gly  
115 120 125  
40 Xaa Leu Ser Leu Met Leu Ile Cys Thr Leu Leu Gln Phe Cys Leu Ala  
130 135 140  
Val Leu Thr Ala Val Leu Arg Trp Lys Gln Ala Tyr Ser Asp Phe Pro  
145 150 155 160  
45 Gly Ser Val Leu Phe Leu Pro His Ser Tyr Ile Gly Asn Ser Gly Met  
165 170 175  
50 Ser Ser Lys Met Thr His Asp Cys Gly Tyr Gln Gln Leu Leu Thr Ser  
180 185 190

## (2) INFORMATION FOR SEQ ID NO: 471:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 234 amino acids

(B) TYPE: amino acid

60

637

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 471:

5 Met Arg Lys Thr Arg Leu Trp Gly Leu Leu Trp Met Leu Phe Val Ser 15  
1 5 10  
Glu Leu Arg Ala Thr Lys Leu Thr Glu Glu Lys Tyr Glu Leu Lys 30  
20 25  
10 Glu Gly Gln Thr Leu Asp Val Lys Cys Asp Tyr Thr Leu Glu Lys Phe 45  
35 40  
Ala Ser Ser Gln Lys Ala Trp Gln Ile Ile Arg Asp Gly Glu Met Pro 60  
50 55  
15 Lys Thr Leu Ala Cys Thr Glu Arg Pro Ser Lys Asn Ser His Pro Val 75  
65 80  
Gln Val Gly Arg Ile Ile Leu Glu Asp Tyr His Asp His Gly Leu Leu 95  
85 90  
20 Arg Val Arg Met Val Asn Leu Gln Val Glu Asp Ser Gly Leu Tyr Gln 105  
100 110  
25 Cys Val Ile Tyr Gln Pro Pro Lys Glu Pro His Met Leu Phe Asp Arg 115  
120 125  
Ile Arg Leu Val Val Thr Lys Gly Phe Ser Gly Thr Pro Gly Ser Asn 135  
130 140  
30 Glu Asn Ser Thr Gln Asn Val Tyr Lys Ile Pro Pro Thr Thr Thr Lys 145  
150 155 160  
Ala Leu Cys Pro Leu Tyr Thr Ser Pro Arg Thr Val Thr Gln Ala Pro 165  
170 175  
35 Pro Lys Ser Thr Ala Asp Val Ser Thr Pro Asp Ser Glu Ile Asn Leu 180  
185 190  
40 Thr Asn Val Thr Asp Ile Ile Arg Val Pro Val Phe Asn Ile Val Ile 195  
200 205  
Leu Leu Ala Gly Gly Phe Leu Ser Lys Ser Leu Val Phe Ser Val Leu 210  
215 220  
45 Phe Ala Val Thr Leu Arg Ser Phe Val Pro 225  
230  
50 (2) INFORMATION FOR SEQ ID NO: 472:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 105 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 472:  
Met Leu His Ile Leu Pro Leu Lys Ser Tyr Asp Phe Pro His Phe Ser 15  
1 5 10

638

Leu Met Gly Arg Tyr Arg Cys Ala Ser Leu Leu Phe Cys Phe Leu Leu 30  
20 25  
5 Leu Phe Phe Phe Cys Ser Val Leu Trp Thr Phe Ser Asp Met His 35  
40 45  
Arg Ser Gly Glu Asp Gly Pro Trp Thr Pro Cys Val His His Leu Ala 50  
55 60  
10 Ala Ser Leu Ile Ser Tyr Gly Gln Pro Gly Phe Ile Cys Ile Ser Leu 65  
70 75 80  
Phe Ser Pro Val Leu Phe Ile Glu Asn Pro Arg His Tyr Ala Asn Ala 85  
90 95  
Thr Val Thr Thr Leu Gly Asp Trp Xaa 100  
105  
20 (2) INFORMATION FOR SEQ ID NO: 473:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 473:  
Met Val Phe Leu Lys Tyr Arg Phe Leu Phe Phe Leu Val Phe Leu Ala 15  
1 5 10  
30 Asn Cys Ile Tyr Ser Leu His Tyr Lys Pro Ser Leu Met Tyr Pro Lys 20  
25 25 30  
40 (2) INFORMATION FOR SEQ ID NO: 474:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 571 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 474:  
Met Ala Leu Ser Arg Gly Leu Pro Arg Glu Leu Ala Glu Ala Val Ala 15  
1 5 10  
50 Gly Gly Arg Val Leu Val Val Gly Ala Gly Gly Ile Gly Cys Glu Leu 20  
25 30  
55 Leu Lys Asn Leu Val Leu Thr Gly Phe Ser His Ile Asp Leu Ile Asp 35  
40 45  
Leu Asp Thr Ile Asp Val Ser Asn Leu Asn Arg Gln Phe Leu Phe Gln 50  
55 60

Lys Lys His Val Gly Arg Ser Lys Ala Gln Val Ala Lys Glu Ser Val  
65 70 75 80  
5 Leu Gln Phe Tyr Pro Lys Ala Asn Ile Val Ala Tyr His Asp Ser Ile  
85 90 95  
Met Asn Pro Asp Tyr Asn Val Glu Phe Phe Arg Gln Phe Ile Leu Val  
100 105 110  
10 Met Asn Ala Leu Asp Asn Arg Ala Ala Arg Asn His Val Asn Arg Met  
115 120 125  
15 Cys Leu Ala Ala Asp Val Pro Leu Ile Glu Ser Gly Thr Ala Gly Tyr  
130 135 140  
Leu Gly Gln Val Thr Thr Ile Lys Lys Gly Val Thr Glu Cys Tyr Glu  
145 150 155 160  
20 Cys His Pro Lys Pro Thr Gln Arg Thr Phe Pro Gly Cys Thr Ile Arg  
165 170 175  
Asn Thr Pro Ser Glu Pro Ile His Cys Ile Val Trp Ala Lys Tyr Leu  
180 185 190  
25 Phe Asn Gln Leu Phe Gly Glu Glu Asp Ala Asp Gln Glu Val Ser Pro  
195 200 205  
30 Asp Arg Ala Asp Pro Glu Ala Ala Trp Glu Pro Thr Glu Ala Glu Ala  
210 215 220  
Arg Ala Arg Ala Ser Asn Glu Asp Gly Asp Ile Lys Arg Ile Ser Thr  
225 230 235 240  
35 Lys Glu Trp Ala Lys Ser Thr Gly Tyr Asp Pro Val Lys Leu Phe Thr  
245 250 255  
Lys Leu Phe Lys Asp Asp Ile Arg Tyr Leu Leu Thr Met Asp Lys Leu  
260 265 270  
40 Trp Arg Lys Arg Lys Pro Pro Val Pro Leu Asp Trp Ala Glu Val Gln  
275 280 285  
Ser Gln Gly Glu Glu Thr Asn Ala Ser Asp Gln Gln Asn Glu Pro Gln  
290 295 300  
45 Leu Gly Leu Lys Asp Gln Gln Val Leu Asp Val Lys Ser Tyr Ala Arg  
305 310 315 320  
50 Leu Phe Ser Lys Ser Ile Glu Thr Leu Arg Val His Leu Ala Glu Lys  
325 330 335  
Gly Asp Gly Ala Glu Leu Ile Trp Asp Lys Asp Asp Pro Ser Ala Met  
340 345 350  
55 Asp Phe Val Thr Ser Ala Ala Asn Leu Arg Met His Ile Phe Ser Met  
355 360 365  
Asn Met Lys Ser Arg Phe Asp Ile Lys Ser Met Ala Gly Asn Ile Ile  
370 375 380 385 390 395 400

Pro Ala Ile Ala Thr Thr Asn Ala Val Ile Ala Gly Leu Ile Val Leu  
385 390 395 400  
5 Glu Gly Leu Lys Ile Leu Ser Gly Lys Ile Asp Gln Cys Arg Thr Ile  
405 410 415  
Phe Leu Asn Lys Gln Pro Asn Pro Arg Lys Lys Leu Leu Val Pro Cys  
420 425 430  
10 Ala Leu Asp Pro Pro Asn Pro Asn Cys Tyr Val Cys Ala Ser Lys Pro  
435 440 445  
15 Glu Val Thr Val Arg Leu Asn Val His Lys Val Thr Val Leu Thr Leu  
450 455 460  
Gln Asp Lys Ile Val Lys Glu Lys Phe Ala Met Val Ala Pro Asp Val  
465 470 475 480  
20 Gln Ile Glu Asp Gly Lys Gly Thr Ile Leu Ile Ser Ser Glu Glu Gly  
485 490 495  
Glu Thr Glu Ala Asn Asn His Lys Lys Leu Ser Glu Phe Gly Ile Arg  
500 505 510  
25 Asn Gly Ser Arg Leu Gln Ala Asp Asp Phe Leu Gln Asp Tyr Thr Leu  
515 520 525  
30 Leu Ile Asn Ile Leu His Ser Glu Asp Leu Gly Lys Asp Val Glu Phe  
530 535 540  
Glu Val Val Gly Asp Ala Pro Glu Lys Val Gly Xaa Lys Gln Ala Glu  
545 550 555 560  
35 Asp Ala Ala Lys Ser Ile Thr Asn Gly Gln Xaa  
565 570 575  
40 (2) INFORMATION FOR SEQ ID NO: 475:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 312 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 475:  
Met Gln Val Val Thr Cys Leu Thr Arg Asp Ser Tyr Leu Thr His Cys  
1 5 10 15  
50 Phe Leu Gln His Leu Met Val Val Leu Ser Ser Leu Glu Arg Thr Pro  
20 25 30  
Ser Pro Glu Pro Val Asp Lys Asp Phe Tyr Ser Glu Phe Gly Asn Lys  
35 40 45  
55 Thr Thr Gly Lys Met Glu Asn Tyr Glu Leu Ile His Ser Ser Arg Val  
50 55 60  
Lys Phe Thr Tyr Pro Ser Glu Glu Glu Ile Gly Asp Leu Thr Phe Thr  
65 70 75 80 85 90 95 100



Val Ala Gln Lys Met Ala Glu Pro Glu Lys Ala Pro Ala Leu Ser Ile 85 90 95  
5 Leu Leu Tyr Val Gln Ala Phe Gln Val Gln Gly Met Pro Pro Gly Cys 100 105 110  
Cys Arg Gly Pro Leu Arg Pro Lys Thr Leu Leu Thr Ser Ser Glu 115 120 125  
10 Ile Phe Leu Leu Asp Glu Asp Cys Val His Tyr Pro Leu Pro Glu Phe 130 135 140  
Ala Lys Glu Pro Pro Gln Arg Asp Arg Tyr Arg Leu Asp Asp Gly Arg 145 150 155 160  
Arg Val Arg Asp Leu Asp Arg Val Leu Met Gly Tyr Gln Thr Tyr Pro 165 170 175  
20 Gln Pro Ser Pro Ser Ser Met Thr Cys Lys Val Met Thr Ser Trp 180 185 190  
Ala Val Ser Pro Trp Thr Thr Leu Gly Arg Cys Gln Val Ala Arg Leu 195 200 205  
25 Glu Pro Ala Arg Ala Val Lys Ser Ser Gly Arg Cys Leu Ser Pro Val 210 215 220  
Leu Arg Ala Glu Arg Ser Ser Arg Cys Trp Leu Ala Ser Gly Arg 225 230 235 240  
Pro Cys Val Ala Val Ser Cys Leu Ser Ser Ser Pro Ala Ser Pro Gly 245 250 255  
35 His Ser Gln Pro Val Val Ser Ser Ser Leu Thr Pro Thr Gly Ala Gly Gln 260 265 270  
Gln Ala Phe Val Phe Ser Lys Asn Val Leu Ser Ser Leu Trp Tyr Leu 275 280 285  
40 Asn Leu Thr Val Leu Ala Glu Asn Val Asn Met Cys Val Cys Cys Val 290 295 300  
Asn Ser Phe Ser Cys Trp Glu Xaa 305 310  
50 (2) INFORMATION FOR SEQ ID NO: 476:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 329 amino acids  
(B) TYPE: amino acid  
(C) TOPOLOGY: linear  
55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 476:  
Met Ala Gln His His Leu Trp Ile Leu Leu Cys Leu Gln Thr Trp 1 5 10 15  
60 Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile

20 25 30  
Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln 35 40 45  
5 Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr 50 55 60  
Pro Gly Asp Ser Glu Thr Ala Pro Val Val Thr Val Thr His Arg Asn 65 70 75 80  
10 Tyr Tyr Glu Arg Ile His Ala Leu Gly Pro Asn Tyr Asn Leu Val Ile 85 90 95  
15 Ser Asp Leu Arg Met Glu Asp Ala Gly Asp Tyr Lys Ala Asp Ile Asn 100 105 110  
Thr Gln Ala Asp Pro Tyr Thr Thr Lys Arg Tyr Asn Leu Gln Ile 115 120 125  
20 Tyr Arg Arg Leu Gly Lys Pro Lys Ile Thr Gln Ser Leu Met Ala Ser 130 135 140  
Val Asn Ser Thr Cys Asn Val Thr Leu Thr Cys Ser Val Glu Lys Glu 145 150 155 160  
Glu Lys Asn Val Thr Tyr Asn Trp Ser Pro Leu Gly Glu Gly Asn 165 170 175  
30 Val Leu Gln Ile Phe Gln Thr Pro Glu Asp Gln Glu Leu Thr Tyr Thr 180 185 190  
Cys Thr Ala Gln Asn Pro Val Ser Asn Asn Ser Asp Ser Ile Ser Ala 195 200 205  
35 Arg Gln Leu Cys Ala Asp Ile Ala Met Gly Phe Arg Thr His His Thr 210 215 220  
Gly Leu Leu Ser Val Leu Ala Met Phe Phe Leu Leu Val Leu Ile Leu 225 230 235 240  
Ser Ser Val Phe Leu Phe Arg Leu Phe Lys Arg Arg Gln Asp Ala Ala 245 250 255  
40 45 Ser Lys Lys Thr Ile Tyr Thr Tyr Ile Met Ala Ser Arg Asn Thr Gln 260 265 270  
Pro Ala Glu Ser Arg Ile Tyr Asp Glu Ile Leu Gln Ser Lys Val Leu 275 280 285  
50 Pro Ser Lys Glu Glu Pro Val Asn Thr Val Tyr Ser Glu Val Gln Phe 290 295 300  
Ala Asp Lys Met Gly Lys Ala Ser Thr Gln Asp Ser Lys Pro Pro Gly 305 310 315 320  
Thr Ser Ser Tyr Glu Ile Val Ile Xaa 325  
60

## (2) INFORMATION FOR SEQ ID NO: 477:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 178 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 477:

10 Met Lys Leu Gln Cys Val Ser Leu Trp Leu Leu Gly Thr Ile Leu Ile  
1 5 10 15  
Leu Cys Ser Val Asp Asn His Gly Leu Arg Arg Cys Leu Ile Ser Thr  
20 25 30  
Asp Met His His Ile Gln Gln Ser Phe Gln Gln Ile Lys Arg Ala Ile  
35 40 45  
Gln Ala Lys Asp Thr Phe Pro Asn Val Thr Ile Leu Ser Thr Leu Gln  
50 55 60  
Thr Leu Gln Ile Ile Lys Pro Leu Asp Val Cys Cys Val Thr Lys Asn  
65 70 75 80  
Leu Leu Ala Phe Tyr Val Asp Arg Val Phe Lys Asp His Gln Gln Pro  
85 90 95  
Asn Pro Lys Ile Leu Arg Lys Ile Ser Ser Ile Ala Asn Ser Phe Leu  
100 105 110  
Tyr Met Gln Lys Thr Leu Arg Gln Cys Gln Gln Arg Gln Cys His  
115 120 125  
Cys Arg Gln Gln Ala Thr Asn Ala Thr Arg Val Ile His Asp Asn Tyr  
130 135 140  
Asp Gln Leu Gln Val His Ala Ala Ile Lys Ser Leu Gly Gln Leu  
145 150 155 160  
Asp Val Phe Leu Ala Trp Ile Asn Lys Asn His Gln Val Met Ser Ser  
165 170 175  
Ala Xaa

45

## (2) INFORMATION FOR SEQ ID NO: 478:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 478:

55 Asp Thr Ala Ile Arg Val Ala Leu Ala Val Ala Val Leu Lys Thr Val  
1 5 10 15  
Ile Leu Gly Leu Leu Cys Leu Leu Leu Cys Gly Gly Gly Gly Lys  
20 25 30

60

Val Ala Gly Arg Gln Ala Val Thr Ser Asp Gln Gln Ser Val Gly Arg  
35 40 45

5 Arg Asp Val Tyr  
50

## (2) INFORMATION FOR SEQ ID NO: 479:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 62 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 479:

10 Met Gln Lys Lys Asn Ser Leu Phe Phe Phe Ala Phe Tyr Tyr Gln  
1 5 10 15  
Asn Lys Thr Asn Ala Pro Gly Gln Gly Ser Met Ile Thr Arg Asn Ile  
20 25 30  
Lys Gln Tyr Phe Leu Pro Phe Leu Phe Cys Cys Val Gln Ala Ser Ile  
35 40 45  
Ala Ile Asn Lys Leu Asn Tyr Leu His Trp Thr His Phe Gln  
50 55 60

30

## (2) INFORMATION FOR SEQ ID NO: 480:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

40 Met Pro Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser  
1 5 10 15  
Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly  
20 25

45

## (2) INFORMATION FOR SEQ ID NO: 481:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 339 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

55 Met Ser Gly Pro Asp Val Gln Thr Pro Ser Ala Ile Gln Ile Cys Arg  
1 5 10 15  
Ile Met Arg Pro Asp Asp Ala Asn Val Ala Gly Asn Val His Gly Gly  
20 25 30

60

645

Thr Ile Leu Lys Met Ile Glu Glu Ala Gly Ala Ile Ile Ser Thr Arg  
35 40 45

5 His Cys Asn Ser Gln Asn Gly Glu Arg Cys Val Ala Ala Leu Ala Arg  
50 55 60

Val Glu Arg Thr Asp Phe Leu Ser Pro Met Cys Ile Gly Glu Val Ala  
65 70 75 80

10 His Val Ser Ala Glu Ile Thr Tyr Thr Ser Lys His Ser Val Glu Val  
85 90 95

15 Gln Val Asn Val Met Ser Glu Asn Ile Leu Thr Gly Ala Lys Lys Leu  
100 105 110

Thr Asn Lys Ala Thr Leu Trp Tyr Val Pro Leu Ser Leu Lys Asn Val  
115 120 125

20 Asp Lys Val Leu Glu Val Pro Val Val Tyr Ser Arg Xaa Glu Gln  
130 135 140

Glu Glu Glu Gly Arg Lys Arg Tyr Glu Ala Gln Lys Leu Glu Arg Met  
145 150 155 160

25 Glu Thr Lys Trp Arg Asn Gly Asp Ile Val Gln Pro Val Leu Asn Pro  
165 170 175

30 Glu Pro Asn Thr Val Ser Tyr Ser Gln Ser Ser Leu Ile His Leu Val  
180 185 190

Gly Pro Ser Asp Cys Thr Leu His Gly Phe Val His Gly Gly Val Thr  
195 200 205

35 Met Lys Leu Met Asp Glu Val Ala Gly Ile Val Ala Ala Arg His Cys  
210 215 220

Lys Thr Asn Ile Val Thr Ala Ser Val Asp Ala Ile Asn Phe His Asp  
225 230 235 240

40 Lys Ile Arg Lys Gly Cys Val Ile Thr Ile Ser Gly Arg Met Thr Phe  
245 250 255

45 Thr Ser Asn Lys Ser Met Glu Ile Glu Val Leu Val Asp Ala Asp Pro  
260 265 270

Val Val Asp Ser Ser Gln Lys Arg Tyr Arg Ala Ala Ser Ala Phe Phe  
275 280 285

50 Thr Tyr Val Ser Leu Ser Gln Glu Gly Arg Ser Leu Pro Val Pro Gln  
290 295 300

Leu Val Pro Glu Thr Glu Asp Glu Lys Lys Arg Phe Glu Glu Gly Lys  
305 310 315 320

55 Gly Arg Tyr Leu Gln Met Lys Ala Lys Xaa Gln Gly His Ala Xaa Xaa  
325 330 335

60 Gln Pro Xaa

646

(2) INFORMATION FOR SEQ ID NO: 482:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

Met Leu Asn Ser Asn Ile Asn Asp Leu Leu Met Val Thr Tyr Leu Ala  
1 5 10 15

15 Asn Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn Leu  
20 25 30

20

(2) INFORMATION FOR SEQ ID NO: 483:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 483:

Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu  
1 5 10 15

35 Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr Arg Ile Gly  
20 25 30

Cys Phe Lys Thr Ile Thr Cys Trp Pro Thr Ser Leu Thr Gln Arg Xaa  
35 40 45

40

45 (2) INFORMATION FOR SEQ ID NO: 484:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 484:

Met Tyr Met Tyr Ser Leu Asn Val Phe Leu Ser Phe Ile Phe Leu Ala  
1 5 10 15

55 Leu Val Phe Lys Cys Val His Val Cys Gln Gly Ala Asn Ala Phe Leu  
20 25 30

60 Phe Leu Lys Leu Val Phe  
35

## (2) INFORMATION FOR SEQ ID NO: 485:

5

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 485:

10

Met Gly Leu Arg Leu Ile Cys Leu Glu Leu Thr Met Val Lys Ala Leu  
1 5 10 15

15 Val Cys Glu Met Phe Leu Phe Leu Met Thr Gln Lys Leu Ile Thr  
20 25 30

20 Gln Glu Cys Thr Glu Lys Phe Ala Lys Leu Leu Val Gln Leu Ile Ser  
35 40 45

Leu Val Phe Ala Trp Glu Phe Phe Ser Glu Asp Thr Pro  
50 55 60

25

## (2) INFORMATION FOR SEQ ID NO: 486:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 346 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 486:

30

Met Leu Ala Ala Arg Leu Val Cys Leu Arg Thr Leu Pro Ser Arg Val  
1 5 10 15

35 Phe His Pro Ala Phe Thr Lys Ala Ser Pro Val Val Lys Asn Ser Ile  
20 25 30

40 Thr Lys Asn Gln Trp Leu Leu Thr Pro Ser Arg Glu Tyr Ala Thr Lys  
35 40 45

45 Thr Arg Ile Gly Ile Arg Arg Gly Arg Thr Gly Gln Glu Leu Lys Glu  
50 55 60

Ala Ala Leu Glu Pro Ser Met Glu Lys Ile Phe Lys Ile Asp Gln Met  
65 70 75 80

50 Gly Arg Trp Phe Val Ala Gly Gly Ala Ala Val Gly Leu Gly Ala Leu  
85 90 95

Cys Tyr Tyr Gly Leu Gly Leu Ser Asn Glu Ile Gly Ala Ile Glu Lys  
100 105 110

55 Ala Val Ile Trp Pro Gln Tyr Val Lys Asp Arg Ile His Ser Thr Tyr  
115 120 125

60 Met Tyr Leu Ala Gly Ser Ile Gly Leu Thr Ala Leu Ser Ala Ile Ala  
130 135 140

Ile Ser Arg Thr Pro Val Leu Met Asn Phe Met Met Arg Gly Ser Trp  
145 150 155 160

5 Val Thr Ile Gly Val Thr Phe Ala Met Val Gly Ala Gly Met Leu  
165 170 175

Val Arg Ser Ile Pro Tyr Asp Gln Ser Pro Gly Pro Lys His Leu Ala  
180 185 190

10 Trp Leu Leu His Ser Gly Val Met Gly Ala Val Val Ala Pro Leu Thr  
195 200 205

Ile Leu Gly Gly Pro Leu Leu Ile Arg Ala Ala Trp Tyr Thr Ala Gly  
210 215 220

15 Ile Val Gly Gly Leu Ser Thr Val Ala Met Cys Ala Pro Ser Glu Lys  
225 230 235 240

20 Phe Leu Asn Met Gly Ala Pro Leu Gly Val Gly Leu Gly Leu Val Phe  
245 250 255

Val Ser Ser Leu Gly Ser Met Phe Leu Pro Pro Thr Thr Val Ala Gly  
260 265 270

25 Ala Thr Leu Tyr Ser Val Ala Met Tyr Gly Gly Leu Val Leu Phe Ser  
275 280 285

Met Phe Leu Leu Tyr Asp Thr Gln Lys Val Ile Lys Arg Ala Glu Val  
290 295 300

30 Ser Pro Met Tyr Gly Val Gln Lys Tyr Asp Pro Ile Asn Ser Met Leu  
305 310 315 320

35 Ser Ile Tyr Met Asp Thr Leu Asn Ile Phe Met Arg Val Ala Thr Met  
325 330 335

Leu Ala Thr Gly Gly Asn Arg Lys Lys Xaa  
340 345

## (2) INFORMATION FOR SEQ ID NO: 487:

40

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 487:

45

Met Glu Glu Val Leu Leu Leu Gly Leu Lys Asp Arg Glu Gly Tyr Thr  
1 5 10 15

Ser Phe Trp Asn Asp Cys Ile Ser Ser Gly Leu Arg Gly Cys Met Leu  
20 25 30

55 Ile Glu Leu Ala Leu Arg Gly Arg Leu Gln Leu Ala Cys Gly Met  
35 40 45

60 Arg Arg Lys Ser Leu Leu Thr Arg Lys Val Ile Cys Lys Ser Asp Ala  
50 55 60

649

Pro Thr Gly Asp Val Leu Leu Asp Glu Ala Leu Lys His Val Lys Glu 80  
65 70 75

5 Thr Gln Pro Pro Glu Thr Val Gln Asn Trp Ile Glu Leu Leu Ser Gly 95  
85 90

Glu Thr Trp Asn Pro Leu Lys Leu His Tyr Gln Leu Arg Asn Val Arg 110  
100 105 110

10 Glu Arg Leu Ala Lys Asn Leu Val Glu Lys Gly Val Leu Thr Thr Glu 125  
115 120 125

Lys Gln Asn Phe Leu Leu Phe Asp Met Thr Thr His Pro Leu Thr Asn 140  
130 135 140

15 Asn Asn Ile Lys Gln Arg Leu Ile Lys Lys Val Gln Glu Ala Val Leu 160  
145 150 155

20 Asp Lys Trp Val Asn Asp Pro His Arg Met Asp Arg Arg Leu Leu Ala 175  
165 170 175

Leu Ile Tyr Leu Ala Ala Ser Asp Val Leu Glu Asn Ala Phe Ala 190  
180 185 190

25 Pro Leu Leu Asp Glu Gln Tyr Asp Leu Ala Thr Lys Arg Val Arg Gln 205  
195 200 205

30 Leu Leu Asp Leu Asp Pro Glu Val Glu Cys Leu Lys Ala Asn Thr Asn 220  
210 215 220

Glu Val Leu Trp Ala Val Val Ala Phe Thr Lys Xaa 235  
225 230 235

35

(2) INFORMATION FOR SEQ ID NO: 488:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 200 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 488:

40 Met Ala Gln Arg Met Val Trp Val Asp Leu Glu Met Thr Thr Gly Leu Asp 15  
1 5 10

Ile Glu Lys Asp Gln Ile Ile Glu Met Ala Cys Leu Ile Thr Asp Ser 30  
20 25 30

50 Asp Leu Asn Ile Leu Ala Glu Gly Pro Asn Leu Ile Ile Lys Gln Pro 45  
35 40 45

55 Asp Glu Leu Leu Asp Ser Met Ser Asp Trp Cys Lys Glu His His Gly 60  
50 55 60

Lys Ser Gly Leu Thr Lys Ala Val Lys Glu Ser Thr Thr Thr Leu Gln 80  
65 70 75

60 Gln Ala Glu Tyr Glu Phe Leu Ser Phe Val Arg Gln Gln Thr Pro Pro

650

85 90 95

Gly Leu Cys Pro Leu Ala Gly Asn Ser Val His Glu Asp Lys Lys Phe 110  
100 105

5 Leu Asp Lys Tyr Met Pro Gln Phe Met Lys His Leu His Tyr Arg Ile 125  
115 120 125

10 Ile Asp Val Ser Thr Val Lys Glu Leu Cys Arg Arg Trp Tyr Pro Glu 140  
130 135 140

Glu Tyr Glu Phe Ala Pro Lys Lys Ala Ala Ser His Arg Ala Leu Asp 160  
145 150 155

15 Asp Ile Ser Glu Ser Ile Lys Glu Leu Gln Phe Tyr Arg Asn Asn Ile 175  
165 170 175

Phe Lys Lys Lys Ile Asp Glu Lys Lys Arg Lys Ile Ile Glu Asn Gly 190  
180 185 190

20 Glu Asn Glu Lys Thr Val Ser Xaa 200  
195

25

(2) INFORMATION FOR SEQ ID NO: 489:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 351 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 489:

35 Met Ala Thr Thr Ala Ala Pro Ala Gly Gly Ala Arg Asn Gly Ala Gly 15  
1 5 10

Pro Glu Trp Gly Gly Phe Glu Glu Asn Ile Gln Gly Gly Gly Ser Ala 30  
20 25 30

40 Val Ile Asp Met Glu Asn Met Asp Asp Thr Ser Gly Ser Ser Phe Glu 45  
35 40 45

Asp Met Gly Glu Leu His Gln Arg Leu Arg Glu Glu Glu Val Asp Ala 60  
50 55 60

45 Asp Ala Ala Asp Ala Ala Ala Glu Glu Asp Gly Glu Phe Leu 80  
65 70 75

Gly Met Lys Gly Phe Lys Gly Gln Leu Ser Arg Gln Val Ala Asp Gln 95  
85 90 95

50 Met Trp Gln Ala Gly Lys Arg Gln Ala Ser Arg Ala Phe Ser Leu Tyr 110  
100 105 110

55 Ala Asn Ile Asp Ile Leu Arg Pro Tyr Phe Asp Val Glu Pro Ala Gln 125  
115 120 125

Val Arg Thr Gly Leu Leu Glu Ser Met Ile Pro Ile Lys Met Val Asn 140  
130 135 140

60

651

Phe Pro Gln Lys Ile Ala Gly Gln Leu Tyr Gly Pro Leu Met Leu Val  
145 150 155 160  
Phe Thr Leu Val Ala Ile Leu Leu His Gly Met Lys Thr Ser Asp Thr  
165 170 175  
Ile Ile Arg Gln Gly Thr Leu Met Gly Thr Ala Ile Gly Thr Cys Phe  
180 185 190  
10 Gly Tyr Trp Leu Gly Val Ser Ser Phe Ile Tyr Phe Leu Ala Tyr Leu  
195 200 205  
Cys Asn Ala Gln Ile Thr Met Leu Gln Met Leu Ala Leu Leu Gly Tyr  
210 215 220  
15 Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr Tyr Asn Ile His  
225 230 235 240  
Leu His Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val Gly Gly Leu Ser  
245 250 255  
20 Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr Val Gly Pro Thr  
260 265 270  
25 Gln Arg Leu Leu Leu Cys Gly Thr Leu Ala Ala Leu His Met Leu Phe  
275 280 285  
Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Val Gln Gly Ile Leu  
290 295 300  
30 Asp Thr Leu Gln Gly Pro Asn Ile Pro Pro Ile Gln Arg Val Pro Arg  
305 310 315 320  
Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro Thr Thr Val Leu  
325 330 335  
35 Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln Ser His Xaa  
340 345 350  
40 (2) INFORMATION FOR SEQ ID NO: 490:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 265 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 490:  
Met Arg Gly Ser Arg Gly Gly Tyr Ala Gly Gln Met Ala Ala Ser Gly  
1 5 10 15  
Glu Ser Gly Thr Ser Gly Gly Gly Ser Thr Gln Ala Ala Phe Met  
20 25 30  
Thr Phe Tyr Ser Gln Val Lys Gln Ile Gln Lys Arg Asp Ser Val Leu  
35 40 45  
Thr Ser Lys Asn Gln Ile Gln Arg Leu Thr Arg Pro Gly Ser Ser Tyr  
50 55 60

652

Phe Asn Leu Asn Pro Phe Gln Val Leu Gln Ile Asp Pro Gln Val Thr  
65 70 75 80  
5 Asp Gln Gln Ile Lys Lys Arg Phe Arg Gln Leu Ser Ile Leu Val His  
85 90 95  
Pro Asp Lys Asn Gln Asp Asp Ala Asp Arg Ala Gln Lys Ala Phe Gln  
100 105 110  
10 Ala Val Asp Lys Ala Tyr Lys Leu Leu Leu Asp Gln Gln Lys Lys  
115 120 125  
15 Arg Ala Leu Asp Val Ile Gln Ala Gly Lys Gln Tyr Val Gln His Thr  
130 135 140  
Val Lys Gln Arg Lys Lys Gln Leu Lys Lys Gln Gly Lys Pro Thr Ile  
145 150 155 160  
20 Val Gln Gln Asp Asp Pro Gln Leu Phe Lys Gln Ala Val Tyr Lys Gln  
165 170 175  
Thr Met Lys Leu Phe Ala Gln Leu Gln Ile Lys Arg Lys Gln Arg Gln  
180 185 190  
25 Ala Lys Gln Met His Gln Arg Lys Arg Gln Arg Gln Gln Gln Ile Gln  
195 200 205  
30 Ala Gln Gln Lys Ala Lys Arg Gln Arg Gln Trp Gln Lys Asn Phe Gln  
210 215 220  
Gln Ser Arg Asp Gly Arg Val Asp Ser Tyr Arg Asn Phe Gln Ala Asn  
225 230 235 240  
35 Thr Lys Gly Lys Lys Gln Lys Lys Asn Arg Thr Phe Leu Arg Pro Pro  
245 250 255  
Lys Val Lys Met Gln Gln Arg Gln Xaa  
260 265  
40 (2) INFORMATION FOR SEQ ID NO: 491:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 491:  
Asp Ser Met Pro Thr Cys Pro Leu Xaa Ala Ser Leu Gln Cys Gly Pro  
1 5 10 15  
Leu Leu Pro Val Arg Leu Leu Cys Lys Leu  
20 25  
55 60 (2) INFORMATION FOR SEQ ID NO: 492:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 492:  
Leu Leu Pro Val Arg Leu Leu Cys Lys Leu  
20 25  
55 60

653

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 159 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 492:

Met Asn Glu Tyr Arg Val Pro Glu Leu Asn Val Gln Asn Gly Val Leu  
1 5 10 15

Lys Ser Leu Ser Phe Leu Phe Glu Tyr Ile Gly Glu Met Gly Lys Asp  
20 25 30

Tyr Ile Tyr Ala Val Thr Pro Leu Leu Glu Asp Ala Leu Met Asp Arg  
35 40 45

Asp Leu Val His Arg Gln Thr Ala Ser Ala Val Val Gln His Met Ser  
50 55 60

Leu Gly Val Tyr Gly Phe Gly Cys Glu Asp Ser Leu Asn His Leu Leu  
65 70 75 80

Asn Tyr Val Trp Pro Asn Val Phe Glu Thr Ser Pro His Val Ile Gln  
85 90 95

Ala Val Met Gly Ala Leu Glu Gly Leu Arg Val Ala Ile Gly Pro Cys  
100 105 110

Arg Met Leu Gln Tyr Cys Leu Gln Gly Leu Phe His Pro Ala Arg Lys  
115 120 125

Val Arg Asp Val Tyr Trp Lys Ile Tyr Asn Ser Ile Tyr Ile Gly Ser  
130 135 140

Gln Asp Ala Leu Ile Ala His Tyr Pro Arg Ile Tyr Gln Arg Xaa  
145 150 155

## (2) INFORMATION FOR SEQ ID NO: 493:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 279 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 493:

Met Ile Ser Asp Asn Ser Ala Glu Asn Ile Ala Leu Val Thr Ser Met  
1 5 10 15

Tyr Asp Gly Leu Leu Gln Ala Gly Ala Arg Leu Cys Pro Thr Val Gln  
20 25 30

Leu Glu Asp Ile Arg Asn Leu Gln Asp Leu Thr Pro Leu Lys Leu Ala  
35 40 45

Ala Lys Glu Gly Lys Ile Glu Ile Phe Arg His Ile Leu Gln Arg Glu  
50 55 60

Phe Ser Gly Leu Ser His Leu Ser Arg Lys Phe Thr Glu Trp Cys Tyr  
65 70 75 80

654

Gly Pro Val Arg Val Ser Leu Tyr Asp Leu Ala Ser Val Asp Ser Cys  
85 90 95

Glu Glu Asn Ser Val Leu Glu Ile Ile Ala Phe His Cys Lys Ser Pro  
100 105 110

His Arg His Arg Met Val Val Leu Glu Pro Leu Asn Lys Leu Leu Gln  
115 120 125

Ala Lys Trp Asp Leu Ile Pro Lys Phe Phe Leu Asn Phe Leu Cys  
130 135 140

Asn Leu Ile Tyr Met Phe Ile Phe Thr Ala Val Ala Tyr His Gln Pro  
145 150 155 160

Thr Leu Lys Lys Gln Ala Ala Pro His Leu Lys Ala Glu Val Gly Asn  
165 170 175

Ser Met Leu Leu Thr Gly His Ile Leu Ile Leu Leu Gly Gly Ile Tyr  
180 185 190

Leu Leu Val Gly Gln Leu Trp Tyr Phe Trp Arg Arg His Val Phe Ile  
195 200 205

Trp Ile Ser Phe Ile Asp Ser Tyr Phe Glu Ile Leu Phe Leu Phe Gln  
210 215 220

Ala Leu Leu Thr Val Val Ser Gln Val Leu Cys Phe Leu Xaa Ile Glu  
225 230 235 240

Trp Tyr Leu Pro Leu Val Ser Ala Leu Val Leu Gly Trp Leu Asn  
245 250 255

Leu Leu Tyr Tyr Thr Arg Gly Phe Gln His Thr Gly Ile Tyr Ser Val  
260 265 270

Met Ile Gln Lys Pro Trp Xaa  
275

## (2) INFORMATION FOR SEQ ID NO: 494:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 193 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:

Met Ile Arg Cys Gly Leu Ala Cys Glu Arg Cys Arg Trp Ile Leu Pro  
1 5 10 15

Leu Leu Leu Ser Ala Ile Ala Phe Asp Ile Ile Ala Leu Ala Gly  
20 25 30

Arg Gly Trp Leu Gln Ser Ser Asp His Gly Gln Thr Ser Ser Leu Trp  
35 40 45

Trp Lys Cys Ser Gln Glu Gly Gly Ser Gly Ser Tyr Glu Glu Gly  
50 55 60

655

50 55 60  
Cys Gln Ser Leu Met Glu Tyr Ala Trp Gly Arg Ala Ala Ala Met  
65 70 75 80  
Leu Phe Cys Gly Phe Ile Ile Leu Val Ile Cys Phe Ile Leu Ser Phe  
85 90 95  
Phe Ala Leu Cys Gly Pro Gln Met Leu Val Phe Leu Arg Val Ile Gly  
100 105 110  
Gly Leu Leu Ala Leu Ala Val Phe Gln Ile Ile Ser Leu Val Ile  
115 120 125  
Tyr Pro Val Lys Tyr Thr Gln Thr Phe Thr Leu His Ala Asn Xaa Ala  
130 135 140  
Val Thr Tyr Ile Tyr Asn Trp Ala Tyr Gly Phe Gly Trp Ala Ala Thr  
145 150 155 160  
Ile Ile Leu Ile Gly Cys Ala Phe Phe Cys Cys Leu Pro Asn Tyr  
165 170 175  
Glu Asp Asp Leu Leu Gly Asn Ala Lys Pro Arg Tyr Phe Tyr Thr Ser  
180 185 190  
Ala

25  
30  
35  
40  
45  
50  
55  
60

(2) INFORMATION FOR SEQ ID NO: 495:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 205 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 495:

Met Ala Ala Gly Asp Gln Val Phe Ser Gly Ala Gly His Val Xaa Glu  
1 5 10 15  
His Val Ala Gly Gly Arg His Ala Trp Leu Leu Thr Trp Gln Ser Ala  
20 25 30  
Cys Pro Ala Asn Arg Leu Ser Leu Val Pro Leu Val Pro Ser Ala Ser  
35 40 45  
Met Thr Arg Leu Met Arg Xaa Arg Thr Ala Ser Gly Ser Ser Val Ile  
50 55 60  
Leu Trp Met Ala Pro Ala Ala Pro Thr Pro Ala Arg Ala Pro Glu  
65 70 75 80  
Ala Ala Pro Thr Pro Ala Arg Ala Pro Ala Ala Arg Thr Pro Ala  
85 90 95  
Arg Gly Pro Thr Trp Ser Pro Pro Thr Arg Val Leu Leu Gly Thr  
100 105 110

656

Xaa Pro Gly Pro Ser Pro Trp Arg Ser Pro Ala Arg Arg Pro Ala Gln  
115 120 125  
Leu Pro Pro Asp Ser Asp Leu Cys Ser Gly Pro Leu Leu Pro Gly  
130 135 140  
Pro Phe Ser Pro Pro Ala Cys His Thr Ala Pro Asn Ser Val Leu Ile  
145 150 155 160  
Gln Ser Leu Phe Cys Lys Ser Glu Leu Trp Trp Arg Gln Met Arg Ser  
165 170 175  
Ile Thr Trp Val Pro Ser Pro Lys Ala Gly Trp Arg Trp Thr Lys Gly  
180 185 190  
Arg Lys Gln Ala Ser Pro His Arg Ile Leu Phe His Xaa  
195 200 205

20  
25  
30  
35  
40  
45  
50  
55  
60

(2) INFORMATION FOR SEQ ID NO: 496:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 147 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 496:

Met Ala Leu Thr Leu Leu Pro Ser Val Ser Arg Leu Pro Gly Glu Arg  
1 5 10 15  
Met Ala Ala Ser Gly Leu Pro Tyr Val Leu His His Lys Ser Ser Leu  
20 25 30  
Met Lys Val Ile Phe Phe Pro Tyr Pro Val Leu Pro Leu Pro Ala Pro  
35 40 45  
Asn Gly Thr Trp Val Pro Arg Leu Val Leu Gly Leu Gly Ser Gly Asp  
50 55 60  
Gln Val His Tyr Leu Pro Ile Ser Ser Ser Ile Val Asn Tyr Gly Thr  
65 70 75 80  
Ser Val Ser Gly Lys Ser Trp Val Phe Leu Val Tyr Pro Leu His Pro  
85 90 95  
Thr Pro Thr Trp Ser Thr Arg Cys Phe Gln Val Trp Asp Leu Leu Ser  
100 105 110  
Val Glu Leu Pro Asp Lys Gly Glu Gly Asn Thr Arg Arg Ala Ser Gly  
115 120 125  
Val Pro Gly Leu Ser Gln Leu Pro Thr Ser His Lys Pro Ile Lys Gln  
130 135 140  
Glu Tyr Xaa  
145



657

## (2) INFORMATION FOR SEQ ID NO: 497:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 497:

10 Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val  
1 5 10 15

Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys Glu Val Gly  
20 25 30

15 Ala Ala Leu Pro Pro Arg Gly Pro Ser Leu Ser Asp Cys Leu Gly Leu  
35 40 45

20 Pro Pro Trp Thr Pro Trp Gly Pro Ala Trp Thr Leu Ala Gln Ser Xaa  
50 55 60

## (2) INFORMATION FOR SEQ ID NO: 498:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 94 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 498:

35 Met Ser Thr Gly Ala Leu Asn Thr Ser Pro Pro Ala Ser Asn Arg Leu  
1 5 10 15

Glu Ser Thr Leu Asn Glu Tyr Leu Ile Gln Pro Gln Leu His Cys Ser  
20 25 30

40 Ser Val Gln Arg Leu Thr Leu Lys Trp Gly Cys Ser Ser Leu Gln Arg  
35 40 45

45 Asp Gly Gln Ala Val Pro Trp Gly Leu Trp Gln Arg Ala Tyr Pro Ser  
50 55 60

Leu Leu Pro Thr Leu Pro Ser Asp Leu Leu Arg Pro His Ala Val Thr  
65 70 75 80

50 Pro Ser Val Ser Val Ser Val His Thr Cys Glu Ser Ser Xaa  
85 90

## (2) INFORMATION FOR SEQ ID NO: 499:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 499:

658

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Xaa Leu Phe Ser Ser Ser  
1 5 10 15

5 Leu Pro Phe Leu Trp Leu  
20

## (2) INFORMATION FOR SEQ ID NO: 500:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 500:

20 Arg Gly Gly Leu Cys Pro Leu Leu Val Pro Gly Pro Leu Ala Arg Gln  
1 5 10 15

Glu Pro Ser Pro Ser Leu Gln Gly Cys Ser Glu Ser Pro Val Gly Met  
20 25 30

25 Asp

## (2) INFORMATION FOR SEQ ID NO: 501:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 501:

Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu  
1 5 10 15

40 Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser  
20 25

## (2) INFORMATION FOR SEQ ID NO: 502:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 502:

Pro Gly Lys Pro Gln Ala Cys Pro Glu Thr Ser Val Leu Pro  
1 5 10 15

## (2) INFORMATION FOR SEQ ID NO: 503:

## (i) SEQUENCE CHARACTERISTICS:



661

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 508:

5 Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val 15  
1 5 10  
Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala 30  
20 25  
Val Xaa Lys Lys 35

(2) INFORMATION FOR SEQ ID NO: 509:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 509:

15 Met Ala Leu Val Ala Leu Phe Thr Gln Leu Met Arg Xaa Leu Gly Arg 15  
1 5 10  
Cys Pro Gln

(2) INFORMATION FOR SEQ ID NO: 510:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 510:

35 Met Thr Phe Pro Phe Glu Lys Glu Asn Ser Cys Phe Gln Cys Leu Leu 15  
1 5 10  
Phe Asp Ser Trp Arg Glu Gln Thr Arg Thr Asn Ile Gln Pro Gln Arg 30  
20 25

(2) INFORMATION FOR SEQ ID NO: 511:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 511:

50 Met His Leu Leu Asp Phe Phe Arg Asp Leu Val Leu Leu Val Leu Leu 15  
1 5 10  
Met His Leu Leu Asp Phe Phe Arg Asp Leu Val Leu Leu Val Leu Leu 15  
1 5 10

662

Ala Leu Leu Asp Ser Phe Trp Leu Glu Val Gln Lys 25  
20

(2) INFORMATION FOR SEQ ID NO: 512:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 512:

5 Met Cys Leu Ile His Phe Ile Lys Ile Ile Leu Val Phe Ile Leu Lys 15  
1 5 10  
Leu Trp Leu Tyr Ser Gln Lys Cys Pro Lys 25  
20

(2) INFORMATION FOR SEQ ID NO: 513:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 513:

25 Met Ile His Val His Glu Trp Asn Asp Gln Met Leu Met Val Tyr Ile 15  
1 5 10  
Phe Leu Tyr Pro Val Ser Ile Thr Phe Leu Asn Leu Cys Ser Leu Thr 30  
25  
Cys

(2) INFORMATION FOR SEQ ID NO: 514:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

50 Leu Asn Glu Ser Tyr Val Ser Arg Ala Gly Gly Trp Phe Ser Met Phe 15  
1 5 10  
Xaa Leu Ile Phe Phe Leu Leu Ala Leu Gly Ser Xaa Leu Cys Leu Leu 30  
25  
Leu Cys Leu Pro Ser Phe Asn Lys Thr Arg Arg Lys Gln Lys Pro 45  
35 40 45

663

## (2) INFORMATION FOR SEQ ID NO: 515:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 515:

10 1 Ser Ser Lys Thr Pro Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly Ser 15  
5 10  
Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu Gly 20  
25  
15 Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp 35  
40

## (2) INFORMATION FOR SEQ ID NO: 516:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 516:

30 1 Leu Asn Trp

## (2) INFORMATION FOR SEQ ID NO: 517:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 517:

40 1 Phe Ala Phe Cys Ala Glu Leu Met Ile Gln Asn Trp Thr Leu Gly Ala 10  
5 15  
Val Asp Ser Gln Met Asp Asp Met Asp Met Asp Leu Asp Lys Glu Phe 20  
25  
45 Leu Gln Asp Leu Lys Glu Leu Lys Val Leu Val Ala Asp Lys Asp Leu 35  
40  
50 Leu Asp Leu His Lys Ser Leu Val Cys Thr Ala Leu Arg Gly Lys Leu 50  
55  
Gly Val Phe Ser Glu Met Glu Ala Asn Phe Lys Asn Leu Ser Arg Gly 65  
70  
Leu Val Asn Val Ala Ala Lys Leu Thr His Asn Lys Asp Val Arg Asp 85  
90  
60 Leu Phe Val Asp Leu Val Glu Lys Phe Val Glu Pro Cys Arg Ser Asp 100  
105  
110

664

His Trp Pro Leu Ser Asp Val Arg Phe Phe Leu Asn Gln Tyr Ser Ala 115  
120  
125

5 Ser Val His Ser Leu Asp Gly Phe Arg His Gln Ala Ser Gly Thr Ala 130  
135  
140

10 Thr Trp Ala Pro Ser Ala Ala Ala Ser Cys Ala Cys Ile Met Thr Glu 145  
150  
155  
160

Val Pro Asn Ala Pro Pro Thr Leu Thr Ile Lys Leu Leu 165  
170

## (2) INFORMATION FOR SEQ ID NO: 518:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 518:

25 1 Met Trp Lys Asn Leu Gly Ser Gly Ser Val Phe Val Thr Trp Phe Ser 5  
10  
Leu Val Met Ile Leu Ser Gly Ile Gly Pro Leu Gly Asp Ala Glu Asp 20  
25  
30 Ser Ile Ser Asp Val Ser His Arg Leu Arg Pro 35  
40

## (2) INFORMATION FOR SEQ ID NO: 519:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 519:

40 1 Phe Gln Phe Pro Leu Leu Thr Ile Ala Leu Gln Phe Leu 5  
10

## (2) INFORMATION FOR SEQ ID NO: 520:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 520:

55 1 Met His Tyr Val Ile Val Leu Ser Leu Phe Val Val Leu Glu Lys Lys 5  
10  
Asn Lys Met Gly Ser Asp Gly Cys Leu Arg Lys Asn Gly Ser 20  
25  
60

665

666

5

(2) INFORMATION FOR SEQ ID NO: 521:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 521:

Met Ser Arg Ser Ile Val Leu Arg Gly Ser Leu Phe Leu Phe Ser  
1 5 10 15

15

His Tyr Thr Leu Lys Leu Ser Val Ile Lys Gln Thr Asn Arg Lys  
20 25 30

Ile Val Trp Glu Lys Pro Cys Ile Arg Leu Phe Tyr Xaa Val Leu  
35 40 45

20

(2) INFORMATION FOR SEQ ID NO: 522:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:

Met Pro Leu Pro Val Leu Leu Cys Leu Thr Leu Pro Met Pro Leu Pro  
1 5 10 15

30

Ser Ala Thr Ala Arg Gly Gly Asn Arg Thr  
20 25

35

(2) INFORMATION FOR SEQ ID NO: 523:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 58 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 523:

Ser Ser Ile Pro Val Ser Ile Leu Ile Gly Met Lys Leu Ile Leu Tyr  
1 5 10 15

50

Leu Leu Ile Thr Glu Ser Gly Ser His Glu Lys Lys Ser Phe Tyr Pro  
20 25 30

Ser Phe Lys Tyr Met Phe Lys Ile Ile Tyr Val Ser Ala Tyr Cys  
35 40 45

55

Arg Thr Ala Leu Arg Ala Thr Val Ser His  
50 55

60

(2) INFORMATION FOR SEQ ID NO: 524:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 524:

Asn Arg Thr Leu Leu Phe Leu Ile Leu Phe Val Leu Phe Gly Leu Gly  
1 5 10 15

Tyr Gly Phe

(2) INFORMATION FOR SEQ ID NO: 525:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 525:

Met Phe Leu Leu Val Leu Ser Val Phe Cys Asp Phe Met Cys Ser Ile  
1 5 10 15

Ala Pro Arg Cys His Ala Leu Ser Leu Val Ser Leu Arg Ala Gln His  
20 25 30

Leu Ser Leu Phe Ile Thr Cys His  
35 40

35

(2) INFORMATION FOR SEQ ID NO: 526:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 526:

Met Leu Leu Phe Ile Leu Leu Thr Leu Ser Ser Gly Cys Arg Leu Leu  
1 5 10 15

Val Ser Ser Trp Lys Thr Phe Leu Pro His Phe Ser Leu Pro Gly Pro  
20 25 30

50

Arg Glu His Pro Glu Gly Ser Arg Thr Trp Phe Arg Tyr Trp Glu  
35 40 45

Pro Gly Ala His Cys Leu His Cys Ala  
50 55

(2) INFORMATION FOR SEQ ID NO: 527:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 527:  
5 Ala Arg Leu Leu Leu Phe Leu Ser Ser Val His Pro Ser Ile Met Pro  
1 5 10 15  
Ser Cys Asn Gln Leu  
20  
10 (2) INFORMATION FOR SEQ ID NO: 528:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 528:  
20 Met Ser Leu Thr Ser Ser Leu Thr Phe Leu Ser His Ile Leu Leu Leu  
1 5 10 15  
Pro Gln Lys Leu Gln Phe Leu Ser Trp Met Glu Arg Gln Arg Cys  
25 20 25  
Thr Gly Val Ala Lys Tyr Ala  
35  
30 (2) INFORMATION FOR SEQ ID NO: 529:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 128 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 529:  
40 Met Val Leu Arg Leu Ile Gln Leu Ile Phe Leu Ile Phe Phe Ile His  
1 5 10 15  
Ile Ile Ile Leu Leu Ile Pro Gly Ser Arg Pro Cys Gly Ser Trp Val  
45 20 25 30  
Asn Asp Arg Xaa Leu Gly Leu Arg Asp Val Thr His Leu Ile Tyr Leu  
35 40 45  
His Trp Val His Gly His Leu Pro Trp Cys His Pro Tyr Ile Gln Val  
50 55 60  
Glu Phe Ser Ala Leu Ile Glu Ser Thr Ala Gln Leu Gly Leu Pro Phe  
65 70 75 80  
Ser Trp Val Arg Val Ile His Pro Phe Leu Val Leu Pro Cys Leu Tyr  
85 90 95  
Ser Pro Gly Leu Lys Asn(Gly Ile Phe Leu Phe Leu Leu Arg Ala Met  
100 105 110

Pro Gly Gly Met Phe Pro Gly Asn Leu Glu Ala Phe Arg Val Pro Val  
115 120 125  
5 (2) INFORMATION FOR SEQ ID NO: 530:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 82 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 530:  
10 Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser  
1 5 10 15  
Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe  
20 25 30  
Trp Gly Phe Gly Xaa Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr  
25 35 40 45  
Pro Pro Asp Gln Glu Pro Gln Lys Glu Thr Glu Pro Ala Thr Ser  
50 55 60  
His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln  
65 70 75 80  
Pro Asn  
30  
35 (2) INFORMATION FOR SEQ ID NO: 531:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 531:  
45 Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala  
1 5 10 15  
Tyr Trp Thr Met  
20  
50 (2) INFORMATION FOR SEQ ID NO: 532:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 75 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 532:  
55  
60

Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Gln Leu Lys Ile  
1 5 10 15

5 Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu Lys Ile Ile  
20 25 30

Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Trp Ala Ile  
35 40 45

10 Lys Ala Gln Leu Lys Ile Glu Asn Lys Asp Leu Asp Asn Lys Thr Ala  
50 55 60

15 Lys Gly Gly Gly Gln Glu Ala Leu Thr Cys Thr  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 533:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 533:

25 Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser  
1 5 10 15

30 Phe Ile Val Thr Pro Thr Ser Thr Ser Asn Thr Xaa Ser Tyr Leu Trp  
20 25 30

Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Xaa Trp Ala Cys  
35 40 45

35 Thr Leu Asn Ala Val Thr Arg Gly Gly Leu Pro Glu  
50 55 60

(2) INFORMATION FOR SEQ ID NO: 534:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 534:

45 Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe  
1 5 10 15

50 Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg  
20 25 30

55 Ile Leu Phe Phe Ile Val Phe  
35

(2) INFORMATION FOR SEQ ID NO: 535:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 535:

Met Leu  
1 10

(2) INFORMATION FOR SEQ ID NO: 536:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 536:

20 Met Asp Gln Phe Lys Ile Phe Tyr Phe Leu Lys Ala Phe Phe Ala Cys  
1 5 10 15

25 Cys Asn Val Gln Asp Pro Ser Phe Met Gly Glu Thr Gly Ser Tyr  
20 25 30

Leu Asn Ile Gly  
35

(2) INFORMATION FOR SEQ ID NO: 537:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 537:

40 Met Phe Asp Phe Leu Ser Tyr Phe Lys Asp Leu Leu Ser Cys  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 538:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 538:

Met Gly Phe Gly Phe Val Leu Asn Ile Phe Ser Phe Phe Leu Xaa Pro  
1 5 10 15

Pro Leu

60

671

(2) INFORMATION FOR SEQ ID NO: 539:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 539:

10 Leu Leu Leu Trp Thr Leu Leu Ala Xaa Tyr Xaa  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 540:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 108 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 540:

25 Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu  
1 5 10 15

30 Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg  
20 25 30

35 Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe  
35 40 45

40 Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Leu Cys  
50 55 60

35 Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe  
65 70 75 80

Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu  
85 90 95

40 Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Xaa  
100 105

(2) INFORMATION FOR SEQ ID NO: 541:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 106 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 541:

50 Phe Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met  
1 5 10 15

55 Leu Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Tyr Ile Leu Tyr  
20 25 30

60 Leu Arg Thr Leu Glu Ser Lys Leu Val Leu Phe Gly Arg Glu Val Ser  
35 40 45

672

Pro Ala His Gln Tyr Ala Leu Ala Gly Gly Ile Ser Phe Pro Phe Phe  
50 55 60

5 Trp Leu Ala Gly Ala Gly Ser Ala Val Phe Trp Val Leu Gly Ala Thr  
65 70 75 80

Leu Val Val Ile Gly Ser His Ala Ala Phe His Gln Ile Glu Ala Val  
85 90 95

10 Asp Gly Glu Glu Leu Gln Met Glu Pro Val  
100 105

(2) INFORMATION FOR SEQ ID NO: 542:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 136 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 542:

25 Met Asp Arg Phe Thr Val Ala Gly Val Leu Pro Asp Ile Glu Gln Phe  
1 5 10 15

Phe Asn Ile Gly Asp Ser Ser Ser Gly Leu Ile Gln Thr Val Phe Ile  
20 25 30

30 Ser Ser Tyr Met Val Leu Ala Pro Val Phe Gly Tyr Leu Gly Asp Arg  
35 40 45

Tyr Asn Arg Lys Tyr Leu Met Cys Gly Gly Ile Ala Phe Trp Ser Leu  
50 55 60

35 Val Thr Leu Gly Ser Ser Phe Ile Pro Gly Glu His Phe Trp Leu Leu  
65 70 75 80

40 Leu Leu Thr Arg Gly Leu Val Gly Val Gly Glu Ala Ser Tyr Ser Thr  
85 90 95

Ile Ala Pro Thr Leu Ile Ala Asp Leu Phe Val Ala Asp Gln Arg Thr  
100 105 110

45 Gly Cys Ser Ala Ser Ser Thr Leu Pro Phe Arg Trp Ala Val Val Trp  
115 120 125

50 Ala Thr Leu Gln Ala Pro Lys Xaa  
130 135

(2) INFORMATION FOR SEQ ID NO: 543:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 424 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 543:

60



673

674

Met Ala Gly Asp Trp His Trp Ala Leu Arg Val Thr Pro Gly Leu Gly  
1 5 10 15

Val Val Ala Val Leu Leu Leu Phe Leu Val Val Arg Gly Pro Pro Arg  
5 20 25 30

Gly Ala Val Glu Arg His Ser Asp Leu Pro Pro Leu Asn Pro Thr Ser  
35 40 45

Trp Trp Ala Asp Leu Arg Ala Leu Ala Arg Asn Pro Ser Phe Val Leu  
50 55 60

Ser Ser Leu Gly Phe Thr Ala Val Ala Phe Val Thr Gly Ser Leu Ala  
65 70 75 80

Leu Trp Ala Pro Ala Phe Leu Leu Arg Ser Arg Val Val Leu Gly Glu  
85 90 95

Thr Pro Pro Cys Leu Pro Gly Asp Ser Cys Ser Ser Asp Ser Leu  
100 105 110

Ile Phe Gly Leu Ile Thr Cys Leu Thr Gly Val Leu Gly Val Gly Leu  
115 120 125

Gly Val Glu Ile Ser Arg Arg Xaa Arg His Ser Asn Pro Arg Ala Asp  
130 135 140

Pro Leu Val Cys Ala Thr Gly Leu Leu Gly Ser Ala Pro Phe Leu Phe  
145 150 155 160

Leu Ser Leu Ala Cys Ala Arg Gly Ser Ile Val Ala Thr Tyr Ile Phe  
165 170 175

Ile Phe Ile Gly Glu Thr Leu Leu Ser Met Asn Trp Ala Ile Val Ala  
180 185 190

Asp Ile Leu Leu Tyr Val Val Ile Pro Thr Arg Arg Ser Thr Ala Glu  
195 200 205

Ala Phe Gln Ile Val Leu Ser His Leu Leu Gly Asp Ala Gly Ser Pro  
210 215 220

Tyr Leu Ile Gly Leu Ile Ser Asp Arg Leu Arg Arg Asn Trp Pro Pro  
225 230 235 240

Ser Phe Leu Ser Glu Phe Arg Ala Leu Gln Phe Ser Leu Met Leu Cys  
245 250 255

Ala Phe Val Gly Ala Leu Gly Gly Ala Leu Ser Trp Ala Pro Xaa Ser  
260 265 270

Ser Leu Arg Pro Thr Ala Gly Gly His Ser Cys Thr Cys Arg Ala Cys  
275 280 285

Cys Thr Lys Gln Gly Pro Gln Thr Thr Gly Leu Trp Cys Pro Ser Gly  
290 295 300

Ala Ala Pro Ala Cys Pro Trp Pro Val Cys Ser Ser Glu Arg Leu  
305 310 315 320

Pro Leu Thr Tyr Leu His Ile Cys His Ser Xaa Pro Trp Ala His Pro  
325 330 335

Thr Lys Gly Leu Gly Leu Thr Pro Trp Pro Gly Pro Ala Ser Arg Gly  
340 345 350

Thr Leu Gly Arg Val Pro Ala Pro Arg His Tyr Xaa Gly Ser Ser Gly  
355 360 365

Glu Glu Val Gly Val Gln Glu Gly Asp Pro Ser Pro Gln Gly Xaa Pro  
370 375 380

Gln Gly Leu Gly Ala Ile Cys Asn Gly Ile Lys Phe Val Ala Arg Pro  
385 390 395 400

Gln Val Pro Ala Leu Val Phe Leu Trp Val Ala Ser Asp Leu Ala Pro  
405 410 415

Arg Leu His Pro Arg Ala Pro Glu  
420

(2) INFORMATION FOR SEQ ID NO: 544:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 544:

Met Phe Arg Phe Val Ile Cys Leu Phe Leu Trp Leu Val Leu Cys Arg  
1 5 10 15

Asp Ser Thr Ser Ala Ser Arg Ile Ala Leu Tyr Tyr Arg Ile Val Phe  
20 25 30

Leu Ile His Gln Cys Ser Ser  
35

(2) INFORMATION FOR SEQ ID NO: 545:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 545:

Met Leu Pro Trp Xaa Ala Gln Leu Leu Asp Arg Thr Ile Gly Pro Leu  
1 5 10 15

Tyr Leu Leu Phe Val Gln Phe Ser Pro Ala Phe Ser Arg Thr Ser Pro  
20 25 30

Trp Arg Ser Pro Lys Asn Phe Arg Arg Leu Tyr Pro Cys Thr Thr  
35 40 45

Ser Gly Cys Ala Ala Arg Trp Leu Phe Ser  
60

675

50 55

## 5 (2) INFORMATION FOR SEQ ID NO: 546:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 546:

Met Gly Leu Ser Val Leu Leu Pro Leu Cys Leu Leu Gly Pro Gly Arg  
1 5 10 15

Phe Thr Ser Gly Gln Lys Pro Leu Asp Thr Pro Gly Leu Gly Val Pro  
20 25 30

20 Phe

## 25 (2) INFORMATION FOR SEQ ID NO: 547:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 367 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 547:

Met Ala Lys Pro Gln Val Val Val Ala Pro Val Leu Met Ser Lys Leu  
1 5 10 15

Ser Val Asn Ala Pro Gln Phe Tyr Pro Ser Gly Tyr Ser Ser Tyr  
20 25 30

Thr Gln Ser Tyr Gln Asp Gly Cys Gln Asp Tyr Pro Thr Leu Ser Gln  
35 40 45

Tyr Val Gln Asp Phe Leu Asn His Leu Thr Gln Gln Pro Gly Ser Phe  
50 55 60

Gln Thr Gln Ile Gln Gln Phe Ala Gln Thr Leu Asn Gly Cys Val Thr  
65 70 75 80

Thr Asp Asp Ala Leu Gln Gln Leu Val Gln Leu Ile Tyr Gln Gln Ala  
85 90 95

Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr  
100 105 110

Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln  
115 120 125

Leu Leu Leu Gln Arg Cys Arg Thr Gln Tyr Gln Val Lys Asp Gln Ala  
130 135 140

Ala Lys Gly Asp Gln Val Thr Arg Lys Arg Phe His Ala Phe Val Leu  
145 150 155 160

676

Phe Leu Gly Gln Leu Tyr Leu Asn Leu Gln Ile Lys Gly Thr Asn Gly  
165 170 175

Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Gln Leu Leu  
180 185 190

Asn Ala Leu Phe Ser Asn Pro Met Asp Asp Asn Leu Ile Cys Ala Val  
195 200 205

Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Gln Asp Ala Trp Lys Gln  
210 215 220

Lys Gly Lys Met Asp Met Gln Gln Ile Ile Gln Arg Ile Gln Asn Val  
225 230 235 240

Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Leu Lys  
245 250 255

Leu Val Gln Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser  
260 265 270

Thr Tyr Arg Gln Ala Thr Pro Gln Asn Asp Pro Asn Tyr Phe Met Asn  
275 280 285

Gln Pro Thr Phe Tyr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp  
290 295 300

Pro Asp Tyr Gln Gln Lys Tyr Gln Gln Leu Leu Gln Arg Gln Asp Phe  
305 310 315 320

Phe Pro Asp Tyr Gln Gln Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp  
325 330 335

Pro Tyr Leu Asp Asp Ile Asp Asp Gln Met Asp Pro Gln Ile Gln Gln  
340 345 350

Ala Tyr Gln Lys Phe Cys Leu Gln Ser Gln Arg Lys Arg Lys Gln  
355 360 365

## (2) INFORMATION FOR SEQ ID NO: 548:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 548:

Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met  
1 5 10 15

Leu Ile Val Ser Val Leu Ala Leu Ile Pro Gln Thr Thr Leu Thr  
20 25 30

Val Gly Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Cys Leu Ala  
35 40 45

Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro  
60

677

678

50 55 60  
Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu 75  
65 70  
5  
10 (2) INFORMATION FOR SEQ ID NO: 549:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 549:  
15 Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser 15  
1 5 10  
His Cys Trp Gly Leu Pro Leu Ala Cys Gly Thr Phe Val Gln Gly His 30  
20 20 25 30  
Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala 45  
35 40 45  
25 (2) INFORMATION FOR SEQ ID NO: 550:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 168 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 550:  
35 Met Leu Leu Ser Ser Leu Ala Ala Phe Ser Val Ile Ser Val Val Ser Tyr 15  
1 5 10 15  
Leu Ile Leu Ala Leu Leu Ser Val Thr Ile Ser Phe Arg Ile Tyr Lys 30  
20 20 25 30  
Ser Val Ile Gln Ala Val Gln Lys Ser Glu Gly His Pro Phe Lys 45  
35 40 45  
45 Ala Tyr Leu Asp Val Asp Ile Thr Leu Ser Ser Glu Ala Phe His Asn 60  
50 55 60  
Tyr Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile 80  
65 70 75  
50 Ile Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala 95  
85 90  
Val Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile 110  
100 105  
55 Thr Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val 125  
115 120  
60 Tyr Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg 140  
130 135 140

Asp Gln Thr Lys Ser Ile Val Glu Lys Ile Gln Ala Lys Leu Pro Gly 160  
145 150 155  
5 Ile Ala Lys Lys Lys Ala Glu Xaa 165  
165  
10 (2) INFORMATION FOR SEQ ID NO: 551:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 124 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 551:  
15 Ser Val Pro Phe His Leu Leu Val Val Leu Arg Ser Arg Ala Val Arg 15  
1 5 10  
Ala Arg Arg Arg Arg Glu Pro Arg Ser Leu Pro Arg Pro Gly Asp Glu 30  
20 20 25 30  
Glu Leu Gln Leu Leu Cys Gly Ala Arg Ser Asp Phe Leu Glu Arg 45  
35 40 45  
25 Cys Glu Glu Asp Trp Val Cys Leu Trp His His Ala Asp His Ala Ala 60  
50 55  
30 Phe Pro Gly Ser Phe Gln Cys His Gln Cys Gly Phe Leu Pro His Pro 80  
65 70 75  
Gly Ser Ser Leu Cys His His Gln Leu Gln Asp Leu Val Arg His 95  
85 90  
35 Pro Ser Cys Thr Glu Val Arg Arg Arg Pro Ser Ile Gln Ser Leu Pro 110  
100 105  
40 Gly Arg Arg His Tyr Ser Val Leu Arg Ser Phe Pro 120  
115  
45 (2) INFORMATION FOR SEQ ID NO: 552:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 177 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 552:  
50 Met Val His Leu Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp 15  
1 5 10  
55 Val Thr Phe Val Ser Gly Phe Leu Leu Phe Arg Ser Leu Pro Arg His 30  
20 25 30  
Thr Phe Gly Leu Val Gln Ser Lys Leu Phe Pro Phe Tyr Phe His Ile 45  
35 40 45

Ser Met Gly Cys Ala Phe Ile Asn Leu Cys Ile Leu Ala Ser Gln His  
50 55 60  
Ala Trp Ala Gln Leu Thr Phe Trp Glu Ala Ser Gln Leu Tyr Leu Leu  
5 65 70 75 80  
Phe Leu Ser Leu Thr Leu Ala Thr Val Asn Ala Arg Trp Leu Glu Pro  
85 90 95  
10 Arg Thr Thr Ala Met Trp Ala Leu Gln Thr Val Glu Lys Glu Arg  
100 105 110  
Gly Leu Gly Gly Glu Val Pro Gly Ser His Gln Gly Pro Asp Pro Tyr  
115 120 125  
15 Arg Gln Leu Arg Glu Lys Asp Pro Lys Tyr Ser Ala Leu Arg Gln Asn  
130 135 140  
Phe Phe Arg Tyr His Gly Leu Ser Ser Leu Cys Asn Leu Gly Cys Val  
145 150 155 160  
Leu Ser Asn Gly Leu Cys Leu Ala Gly Leu Ala Leu Glu Ile Arg Ser  
165 170  
25 Leu  
30 (2) INFORMATION FOR SEQ ID NO: 553:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 72 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 553:  
Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu  
1 5 10 15  
Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His  
20 25 30  
Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys  
35 40 45  
Thr Glu Gly Ile Pro Cys Ala Lys Ile Pro Glu Trp Val Thr His Leu  
50 55 60  
50 Thr Trp Gln Thr Leu Lys Asn Ser  
65 70  
55 (2) INFORMATION FOR SEQ ID NO: 554:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 554:

Val Leu Arg Ile Ile Cys Leu Trp Pro Cys Gly Thr Thr Leu Pro Leu  
1 5 10 15  
Val Glu Lys Ala His Asp Ser His Ser Ala Asp Pro Val Cys Pro Gly  
20 25 30  
10 Leu Thr Ala His Leu Pro Val Leu Leu Tyr Val Gln Leu  
35 40 45  
15 (2) INFORMATION FOR SEQ ID NO: 555:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 251 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 555:  
Met Lys His Ala Asp Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser  
1 5 10 15  
25 Pro Leu Leu Met Thr Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu  
20 25 30  
Ser Leu Gly Pro Arg Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg  
35 40 45  
30 Gly Phe Met Ile Val Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr  
50 55 60  
Ile Val Tyr Glu Phe Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp  
65 70 75 80  
Arg Cys Asp Pro Gln Asp Cys Thr Leu Gly Gln Cys Pro Ser Val Pro  
85 90 95  
40 Ser Pro Xaa Thr Pro Val Thr Lys Ala Tyr Val Val Arg Thr Glu Gln  
100 105 110  
Gly Thr Gly Pro Pro Leu Pro Thr Thr Ala Ala Leu Gln Gly Pro Arg Leu  
115 120 125  
Trp Phe Leu Thr His Phe Pro Arg Ala Ala Pro Gly Met Trp Pro His  
130 135 140  
50 Cys Cys Leu Pro Leu Gln Ser Trp Gly Leu Lys Gly Leu Tyr Ser Tyr  
145 150 155 160  
Phe Pro Leu Pro Ala Leu Lys Leu Gly Arg Gly Ala Leu Arg Ala Gly  
165 170 175  
55 Pro Thr Lys Gly Leu Val Ala Phe Phe Leu Thr Gln Lys Arg Ser Ala  
180 185 190  
Ile Met Ser Leu Trp Thr Gln Ser His Ser Thr Pro His Thr Glu  
195 200 205  
60

681

Ala Val Ala Ser Gly Pro Lys Val Arg Val Gly Gly Gly Leu Gly Ile  
210 215 220

5 Gln Pro Val Glu Ala Ala Tyr Ser Thr Cys Val Leu Ile Lys Ser Asp  
225 230 235 240

Arg Gly Asn His Glu Lys Lys Lys Lys Lys  
245 250

10

(2) INFORMATION FOR SEQ ID NO: 556:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 556:

20 Gly Leu Ala Gly Leu Cys Gly Gln Leu Ser Ser Pro Ala Leu Cys Val  
1 5 10 15

Asn Arg Leu

25

(2) INFORMATION FOR SEQ ID NO: 557:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 217 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 557:

35 Met Ile Thr Glu Lys Lys Trp Gly Leu Asn Met Glu Tyr Cys Arg Gly Gln  
1 5 10 15

40 Ala Tyr Ile Xaa Ser Ser Gly Phe Ser Ser Lys Met Lys Val Val Ala  
20 25 30

Ser Arg Leu Leu Glu Lys Tyr Pro Gln Ala Ile Tyr Thr Leu Cys Ser  
35 40 45

45 Ser Cys Ala Leu Asn Met Trp Leu Ala Lys Ser Val Pro Val Met Gly  
50 55 60

Val Ser Val Ala Leu Oly Thr Ile Glu Glu Val Cys Ser Phe His  
65 70 75 80

50 Arg Ser Pro Gln Leu Leu Leu Glu Leu Asp Asn Val Ile Ser Val Leu  
85 90 95

55 Phe Gln Asn Ser Lys Glu Arg Gly Lys Glu Leu Lys Glu Ile Cys His  
100 105 110

Ser Gln Trp Thr Gly Arg His Asp Ala Phe Glu Ile Leu Val Glu Leu  
115 120 125

60 Leu Gln Ala Leu Val Leu Cys Leu Asp Gly Ile Asn Ser Asp Thr Asn

682

130 135 140

Ile Arg Trp Asn Asn Tyr Ile Ala Gly Arg Ala Phe Val Leu Cys Ser  
145 150 155 160

5 Ala Val Ser Asp Phe Asp Phe Ile Val Thr Ile Val Val Leu Lys Asn  
165 170 175

10 Val Leu Ser Phe Thr Arg Ala Phe Gly Lys Asn Leu Gln Gly Gln Thr  
180 185 190

Ser Asp Val Phe Phe Ala Ala Gly Ser Leu Thr Ala Val Leu His Ser  
195 200 205

15 Leu Asn Glu Val Ile Gly Lys Tyr Xaa  
210 215

20 (2) INFORMATION FOR SEQ ID NO: 558:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 558:

Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu Asn Glu  
1 5 10 15

30 Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg Asn Thr  
20 25 30

35 Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile Asn Phe  
35 40 45

Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile Lys Leu  
50 55 60

40 Tyr Thr Ser Lys Ser Glu Leu Pro Thr Asp Asn Ser Glu Thr Val Glu  
65 70 75 80

Asn Thr

45

(2) INFORMATION FOR SEQ ID NO: 559:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 559:

Met Val Leu Ile Leu Leu Asn Leu Leu Leu Gly Gln Phe Ser Cys Met  
1 5 10 15

Ser Pro Ala Ser His His Cys His Pro Leu Pro Thr Glu Met Pro Cys  
20 25 30

Ser Ser Asp Trp Gly Phe Asp Ser His Thr Val Tyr Pro Ser Cys Val  
35 40 45  
Asp Ala Leu Leu Pro Lys Pro Ser Ala Asn Ser Phe Pro Asn Gly Ser  
50 55 60  
Cys His Cys Gln Gly Leu Tyr Asn Gln Gln Gln Asn Leu His Ala  
65 70 75 80  
Ala Glu Gly Pro Ala Ser Leu Arg Cys Asn Lys Tyr Val Ser Thr  
85 90 95

15 (2) INFORMATION FOR SEQ ID NO: 560:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 560:

20 Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala Tyr Ala Ile Phe Ile  
1 5 10 15  
Val Phe Thr Val Ile Gly Asp Ala Pro Gly Ala Val Leu Ser Cys Ala  
20 25 30  
Gly His Pro Cys Val Gly Phe Ala Ala Val Leu Val Ala Pro Leu Thr  
35 40 45  
Val Ala Val Ser Ser Xaa  
50

35 (2) INFORMATION FOR SEQ ID NO: 561:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 108 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 561:

40 Met Glu Val Pro Pro Ala Pro Arg Ser Phe Leu Cys Arg Ala Leu  
1 5 10 15  
Cys Leu Phe Pro Arg Val Phe Ala Ala Glu Ala Val Thr Ala Asp Ser  
20 25 30  
Glu Val Leu Glu Glu Arg Gln Lys Arg Leu Pro Tyr Val Pro Glu Pro  
35 40 45  
Tyr Tyr Pro Glu Ser Gly Trp Asp Arg Leu Arg Glu Leu Phe Gly Lys  
50 55 60  
Asp Thr Val Asn Thr Ser Leu Asn Val Tyr Arg Asn Lys Asp Ala Leu  
65 70 75 80

Ser His Phe Val Ile Ala Gly Ala Val Thr Gly Ser Leu Phe Arg Ile  
85 90 95  
Asn Val Gly Leu Arg Gly Trp Trp Leu Val Ala Xaa  
100 105

10 (2) INFORMATION FOR SEQ ID NO: 562:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 50 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 562:

15 Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys Lys  
1 5 10 15  
Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala  
20 25 30  
Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Leu Leu Lys Met Arg Arg  
35 40 45  
Ala Pro  
50

25 (2) INFORMATION FOR SEQ ID NO: 563:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 253 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 563:

30 Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu  
1 5 10 15  
Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu  
20 25 30  
Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp  
35 40 45  
Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr  
50 55 60  
Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu  
65 70 75 80  
Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu  
85 90 95  
Glu Ile Arg Gly Gly Leu Leu Lys Tyr Cys Asn Leu Leu Val Arg  
100 105 110  
Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met  
60

685

686

115

120

125

Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Arg  
130 135 140

Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Leu Glu Asp Arg  
145 150 155 160

Lys Tyr Glu Tyr Leu Met Thr Leu His Gly Val Val Asn Glu Ser Thr  
165 170 175

Val Cys Leu Met Gly His Glu Arg Gln Thr Leu Asn Leu Ile Thr  
180 185 190

Met Leu Ala Ile Arg Val Leu Ala Asp Gln Asn Val Ile Pro Asn Val  
195 200 205

Ala Asn Val Thr Cys Tyr Tyr Gln Pro Ala Pro Tyr Val Ala Asp Ala  
210 215 220

Asn Phe Ser Asn Tyr Tyr Ile Ala Gln Val Gln Pro Val Phe Thr Cys  
225 230 235 240

Gln Gln Gln Thr Tyr Ser Thr Trp Leu Pro Cys Asn Xaa  
245 250

(2) INFORMATION FOR SEQ ID NO: 564:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 564:

Met Ser Phe Leu Met Trp Leu Met Ser Leu Ala Ile Thr Ser Gln Pro  
1 5 10 15

40 Pro Met

(2) INFORMATION FOR SEQ ID NO: 565:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 565:

Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe  
1 5 10 15

Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly  
20 25 30

Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Thr Tyr Ser Ser  
35 40 45

60

Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr  
50 55 60

Gly Ala Ser Tyr His Ser Met Ser Met Ala Arg Ala Ala Phe Phe  
65 70 75 80

10

(2) INFORMATION FOR SEQ ID NO: 566:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 566:

His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val Gln Val Leu Ser  
1 5 10 15

Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu Ala Pro Gly Arg  
20 25 30

Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro Trp Phe Thr Ala  
35 40 45

Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys Arg Gln Arg Arg  
50 55 60

Gln Glu Arg Arg Gln Met Lys Arg Leu  
65 70

(2) INFORMATION FOR SEQ ID NO: 567:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 263 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 567:

Met Asp Cys Pro Ala Leu Pro Gly Trp Lys Lys Glu Glu Val Ile  
1 5 10 15

Arg Lys Ser Gly Leu Ser Ala Gly Lys Ser Asp Val Tyr Tyr Phe Ser  
20 25 30

Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln Leu Ala Arg Tyr Leu  
35 40 45

Gly Asn Thr Val Asp Leu Ser Ser Phe Asp Phe Arg Thr Gly Lys Met  
50 55 60

Met Pro Ser Lys Leu Gln Lys Asn Lys Gln Arg Leu Arg Asn Asp Pro  
65 70 75 80

Leu Asn Gln Asn Lys Gly Lys Pro Asp Leu Asn Thr Thr Leu Pro Ile  
85 90 95  
Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val Thr Lys Val Thr Asn  
100 105 110  
His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln Arg Met Asn Gln Gln  
115 120 125  
Pro Arg Gln Leu Phe Trp Gln Lys Arg Leu Gln Gly Leu Ser Ala Ser  
130 135 140  
Asp Val Thr Gln Gln Ile Ile Lys Thr Met Gln Leu Pro Lys Gly Leu  
145 150 155 160  
Gln Gly Val Gly Pro Gly Ser Asn Asp Gln Thr Leu Leu Ser Ala Val  
165 170 175  
Ala Ser Ala Leu His Thr Ser Ser Ala Pro Ile Thr Gly Gln Val Ser  
180 185 190  
Ala Ala Val Gln Lys Asn Pro Ala Val Trp Leu Asn Thr Ser Gln Pro  
195 200 205  
Leu Cys Lys Ala Phe Ile Val Thr Asp Gln Asp Ile Arg Lys Gln Gln  
210 215 220  
Gln Arg Val Gln Gln Val Arg Lys Lys Leu Gln Gln Ala Leu Met Ala  
225 230 235 240  
Asp Ile Leu Ser Arg Ala Ala Asp Thr Gln Gln Met Asp Ile Gln Met  
245 250 255  
Asp Ser Gly Asp Gln Ala Xaa  
260  
(2) INFORMATION FOR SEQ ID NO: 568:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 568:  
Met Met Arg Pro Phe Tyr Leu Leu Leu Pro Val Leu Cys Thr Gln Ala  
1 5 10 15  
Leu Arg Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Leu  
20 25 30  
Leu Phe Gly Leu Thr His Leu Asn Pro Ser Ala Lys Leu Leu Leu Ser  
35 40 45  
Gln Met Lys Thr Ser Gly Asn Arg Lys Ser Gln Tyr Ser Lys Tyr Ala  
50 55 60  
Arg Asn Trp Lys Lys His  
65 70

(2) INFORMATION FOR SEQ ID NO: 569:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 569:  
Met Pro Val Thr Ser Lys Arg Thr Leu Phe Phe Pro Asp Pro Cys Ser  
1 5 10  
Tyr Asp Thr Pro Pro Pro Asp Cys His Cys His Ser Phe Arg Ala Gln  
20 25 30  
Leu Leu  
20  
(2) INFORMATION FOR SEQ ID NO: 570:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 104 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 570:  
Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu  
1 5 10 15  
Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala  
20 25 30  
Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu  
35 40 45  
His Ala Val Lys Gly Thr Pro Phe Gln Thr Pro Asp Gln Gly Lys Ser  
50 55 60  
Lys Ala Pro Asn Ser Leu Gly Thr Thr Gly Leu Trp Ser Thr Val Tyr  
65 70 75 80  
Ile Phe Thr Gln Val Phe His Asn Phe Ser Asn Asn Ser Ile Phe Ser  
85 90 95  
Gly Lys Phe Leu Tyr Gln Val Xaa  
100

(2) INFORMATION FOR SEQ ID NO: 571:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 132 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 571:



689

690

Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu Leu His Ile Val Leu  
1 10 15

5 Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala Trp Thr Leu Thr Asn  
20 25 30

10 Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu His Ala Val Lys Gly  
35 40 45

Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala Arg Leu Leu Thr His  
50 55 60

15 Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr Ser Arg Lys Phe  
65 70 75 80

Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu Ala Ser Phe Tyr Thr  
85 90 95

20 Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr Ala Ser Leu Leu Ser  
100 105 110

25 Val Leu Ile Pro Lys Met Pro Gln Leu His Gly Val Arg Ile Phe Gly  
115 120 125

Ile Asn Lys Tyr  
130

(2) INFORMATION FOR SEQ ID NO: 572:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 572:

40 Met Asn Lys Trp Ile Cys Glu Met His Cys Tyr Leu Val Leu Leu Ser  
5 10 15

Val Cys Ser Pro Ser Ala Leu Arg Arg Val Arg His Thr Leu Ser Arg  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 573:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 573:

55 Met Pro Val Leu Ser Leu Leu Cys Thr Leu Ile Val Ser Phe Gln Ser  
1 5 10 15

Ala Asp Ser Cys Glu Val Phe Leu Asn Cys Ser Leu  
20 25

(2) INFORMATION FOR SEQ ID NO: 574:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 574:

5 Met Lys Val Ser Thr Met Leu Trp Phe Leu Cys Trp Glu Gln Ser His  
1 5 10 15

15 Phe Leu Arg Glu Trp Glu Asp Leu Ser Thr Phe Leu Ile Leu Ile Gln  
20 25 30

20 Met Glu Cys Gln Tyr Gly Asn Ser  
35 40

(2) INFORMATION FOR SEQ ID NO: 575:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 575:

35 Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser  
1 5 10 15

Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Xaa Leu  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 576:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 576:

45 Met Lys Arg Gly Cys Leu Gly Leu Phe Phe Ser Cys Ser Ser  
1 5 10 15

Ala Pro Thr Met Leu Leu Cys Asp Tyr Leu Asn Trp Phe  
20 25

(2) INFORMATION FOR SEQ ID NO: 577:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

691

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 577:

5 Met Lys Leu Leu Gly Ile Ala Leu Ala Tyr Val Ala Ser Val  
1 5 10 15  
Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu Leu  
20 25 30  
Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu Lys  
35 40 45  
Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val Lys  
50 55 60  
Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Xaa Asn Arg Arg Arg  
65 70 75 80  
Xaa Val Arg Gly Leu Pro Ser Xaa Leu Thr Asn Ser  
85 90

25 (2) INFORMATION FOR SEQ ID NO: 578:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 578:

30 Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg  
1 5 10 15  
Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg  
20 25  
Arg Val Met Val Asn Leu Asn Ile Leu Phe  
35 40

45 (2) INFORMATION FOR SEQ ID NO: 579:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 579:

50 Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Phe Leu Val  
1 5 10 15  
Lys Leu Leu Phe Ile Pro Trp Leu Ala Ser Leu Leu Pro Leu Ser  
20 25  
Thr Phe Thr Ser Asp Phe Tyr Phe Met Glu Phe Gly Ile Glu Val Lys  
35 40 45

60

692

Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr Lys  
50 55 60

5 Lys Phe Asn Lys Lys Lys  
65 70

10 (2) INFORMATION FOR SEQ ID NO: 580:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 110 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 580:

15 Met Leu Arg Leu Leu Leu Val Ala Phe Ala Leu Val Val Val Leu  
1 5 10 15  
Phe His Val Leu Leu Ala Pro Ile Thr Ala Leu Phe His Thr His Phe  
20 25 30  
Tyr Asp Arg Leu Gln Asp Ala Gly Ser Arg Tyr Pro Glu Leu Tyr Leu  
35 40 45  
Tyr Ser Arg Ala Asp Glu Val Val Leu Ala Arg Asp Ile Glu Arg Met  
50 55  
Val Glu Ala Arg Leu Ala Arg Arg Val Leu Ala Arg Ser Val Asp Phe  
65 70 75 80  
Val Ser Ser Ala His Val Ser His Leu Arg Asp Tyr Pro Thr Tyr Tyr  
85 90 95  
Thr Ser Leu Cys Val Asp Phe Met Arg Asn Cys Val Arg Cys  
100 105 110

40 (2) INFORMATION FOR SEQ ID NO: 581:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 581:

45 Met Phe Lys Leu Glu Glu Cys Gly Lys Thr Thr Phe Leu Leu Ser Met  
1 5 10 15  
Ala Leu Tyr Phe Thr Trp Ile Val Gln Thr Thr Lys Gly Cys  
20 25 30

55 (2) INFORMATION FOR SEQ ID NO: 582:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 71 amino acids  
(B) TYPE: amino acid

60

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 582:

Met Glu Ser Asp Ala Leu Leu Thr Ile Phe Trp Ile Ile Ala Arg 15

5

Ser Ser Val Arg Ser Val Gly Lys Ser Ser Gln Arg Ser Phe Thr Thr 30

10

Ile Thr Gln Leu Arg Ser Thr His Thr Gly Pro Ser Arg Arg Ser Tyr 45

Leu Ile Trp Trp Asn Gly Gly Pro Lys Arg Thr Ile Ser Tyr Val Ser 60

15

Arg Arg Phe Arg Ser Phe Arg 70

(2) INFORMATION FOR SEQ ID NO: 583:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 583:

Val Gly Leu Phe Gln Pro Lys Thr Phe Gln Val Pro Val Thr Asp Leu 15

30

Tyr Ile Phe Ile Lys Ile Tyr Ser Glu Ile Gly Pro Ile Met His Val 30

25

Leu Cys Pro Gly Tyr Ser Gln Ser Pro Ser Thr Pro Pro Trp Thr 45

35

(2) INFORMATION FOR SEQ ID NO: 584:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 584:

Met Trp Phe Gly Ser Asp Arg Ser Asp Leu Arg Ile Gly Thr Ala Phe 15

50

Leu Phe Asp Leu Val Cys Asp Leu Cys Ile His Ala Trp Lys Pro Pro 30

Gly Leu Val Arg Phe Ser Phe 35

(2) INFORMATION FOR SEQ ID NO: 585:

60

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 585:

Met Leu Asn Thr Ala Ser Leu Asn Leu Pro Trp Lys Val Gln Leu Phe 15

5

Ala His Ala

(2) INFORMATION FOR SEQ ID NO: 586:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 586:

Met Ser Ala Cys Leu Leu Phe Leu Ala Phe Ser Trp Lys Arg Lys 15

10

Gly Leu Trp Ser Gly Pro Gly 20

(2) INFORMATION FOR SEQ ID NO: 587:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:

Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser 15

40

Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val 30

Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro 45

35

Ser Ser Glu Leu Leu Glu Gln Leu Leu Ser Val Gln Phe Val Trp Gln 60

50

Ala His Thr Val Ala 65

(2) INFORMATION FOR SEQ ID NO: 588:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

60

695

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 588:

Met Gly Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu  
1 5 10 15  
Leu Met Met Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu  
20 25 30  
Ser Thr Cys Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu  
35 40 45  
Asn Val Gly Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr  
50 55 60  
Arg Lys Lys Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys  
65 70 75

20 (2) INFORMATION FOR SEQ ID NO: 589:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 155 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 589:

Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe  
1 5 10 15  
Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30  
Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45  
Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala  
50 55 60  
Val Gly Phe Leu Glu Leu Leu Ala Gly Leu Leu Leu Val Met Gly Pro  
65 70 75  
Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met  
85 90 95  
Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu Ser Thr Cys  
100 105 110  
Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Leu Asn Val Gly  
115 120 125  
Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys  
130 135 140  
Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa  
145 150 155

60

696

(2) INFORMATION FOR SEQ ID NO: 590:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 590:

Met Pro Glu Thr Arg Leu Gly His Arg Gln Gln Phe Ala Val Phe His  
1 5 10 15  
Leu Xaa Pro Val Pro Pro Cys Gly  
20

(2) INFORMATION FOR SEQ ID NO: 591:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 591:

Met Leu Thr Phe Leu Phe Ser Ala Cys Ala Thr Cys Leu Gly Lys Leu  
1 5 10 15  
Ala Ser Pro Leu Ala Pro Val Gly Pro Gln Gln Arg Gly Xaa Pro Pro  
20 25 30  
Gly Pro Pro Leu Leu Ser  
35

35 (2) INFORMATION FOR SEQ ID NO: 592:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 592:

Met Asp Pro Phe His Tyr Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu  
1 5 10 15  
Val Phe Ala Val Val Leu Phe Ser Val Gly Ile Leu Leu Ile Leu Ser  
20 25 30  
Arg Arg Cys Lys Cys Ser Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp  
35 40 45  
Glu Glu Ala Gln Val Glu Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro  
50 55 60  
Gln Lys Ala Glu Asn  
65

60

697

## (2) INFORMATION FOR SEQ ID NO: 593:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 308 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 593:

10 Asn Leu Arg Val Arg Leu Gly Asp Val Ile Ser Ile Gln Pro Cys Pro  
1 5 10 15

Asp Val Lys Tyr Gly Lys Arg Ile His Val Leu Pro Ile Asp Asp Thr  
20 25 30

15 Val Glu Gly Ile Thr Gly Asn Leu Phe Glu Val Tyr Leu Lys Pro Tyr  
35 40 45

Phe Leu Glu Ala Tyr Arg Pro Ile Arg Lys Gly Asp Ile Phe Leu Val  
50 55 60

20 Arg Gly Gly Met Arg Ala Val Glu Phe Lys Val Val Glu Thr Asp Pro  
65 70 75 80

25 Ser Pro Tyr Cys Ile Val Ala Pro Asp Thr Val Ile His Cys Glu Gly  
85 90 95

Glu Pro Ile Lys Arg Glu Asp Glu Glu Ser Leu Asn Glu Val Gly  
100 105 110

30 Tyr Asp Asp Ile Gly Gly Cys Arg Lys Gln Leu Ala Gln Ile Lys Glu  
115 120 125

Met Val Glu Leu Pro Leu Arg His Pro Ala Leu Phe Lys Ala Ile Gly  
130 135 140

35 Val Lys Pro Pro Arg Gly Ile Leu Leu Tyr Gly Pro Pro Gly Thr Gly  
145 150 155 160

Lys Thr Leu Ile Ala Arg Ala Val Ala Asn Glu Thr Gly Ala Phe Phe  
165 170 175

40 Phe Leu Ile Asn Gly Pro Glu Ile Met Ser Lys Leu Ala Gly Glu Ser  
180 185 190

45 Glu Ser Asn Leu Arg Lys Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro  
195 200 205

Ala Ile Ile Phe Ile Asp Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu  
210 215 220

50 Lys Thr His Gly Glu Val Glu Arg Arg Ile Val Ser Gln Leu Leu Thr  
225 230 235 240

Leu Met Asp Gly Leu Lys Gln Arg Ala His Val Ile Val Met Ala Ala  
245 250 255

Thr Asn Arg Pro Asn Ser Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg  
260 265 270

60 Phe Asp Arg Glu Val Asp Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu

698

275 280 285

Glu Ile Leu Gln Ile His Thr Lys Asn Met Lys Leu Ala Asp Asp Val  
290 295 300

5 Asp Leu Glu Gln  
305

## (2) INFORMATION FOR SEQ ID NO: 594:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 594:

10 Met Gln Ile Lys Leu Lys Ser Val Lys Thr Val Phe Ala Ile Thr  
1 5 10 15

20 Leu Leu Val Leu Phe Leu  
20

## (2) INFORMATION FOR SEQ ID NO: 595:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 595:

35 Met Phe Pro Lys Phe Cys Pro Ile Leu Ser Leu Val Asp Phe Ile Ser  
1 5 10 15

40 His Arg Asp Lys Pro Glu Thr Glu  
20

## (2) INFORMATION FOR SEQ ID NO: 596:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 596:

50 Met Leu Ile Glu Cys Ala Trp Gln Leu Met Phe Leu Leu Lys Val  
1 5 10 15

55 Glu Gln Leu Gly Ile Leu Asp Lys  
20

## (2) INFORMATION FOR SEQ ID NO: 597:

699

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 597:

Met  
1

10 (2) INFORMATION FOR SEQ ID NO: 598:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 598:

20 Met Cys Ile Met Ser Ala Leu Val  
1 5

25 (2) INFORMATION FOR SEQ ID NO: 599:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 599:

30 Met Phe Leu Val Tyr Phe Thr Gly Leu Ile Ser Ala Leu Ser Asn  
1 5 10 15

35 Val His Thr Pro Ser Arg Leu Pro Ala  
20 25

40 (2) INFORMATION FOR SEQ ID NO: 600:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 600:

50 Met Xaa Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser  
1 5 10 15

Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly  
20 25

55 (2) INFORMATION FOR SEQ ID NO: 601:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 amino acids

700

- (B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 601:

5 Met Trp Thr Arg Ser Ser Arg Cys Leu Leu Cys Ile Pro Gly Xaa  
1 5 10 15

Ser Arg Arg Arg Ala Gly Ser Gly Met Lys Pro Arg Ser Trp Ser  
20 25 30

10 Ala Trp Arg Pro Ser Gly Gly Thr Gly Thr Ser Ser Gln Ser Ser  
35 40 45

15 Thr Gln Ser Arg Thr Leu Ser Ala Thr Ala Ser Pro Ala  
50 55 60

20 (2) INFORMATION FOR SEQ ID NO: 602:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 602:

25 Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu  
1 5 10 15

30 Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr  
20 25

35 (2) INFORMATION FOR SEQ ID NO: 603:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 603:

40 Met Pro Pro Lys Gln Glu Leu Gly Ser Gly Val Gly Glu Leu Ala Lys  
1 5 10 15

45 Asn Ser Lys Arg Gln His Thr Asn His Arg Trp Lys Tyr Leu Lys  
20 25 30

50 Leu Ile Arg Trp Glu Asp Gly Leu Leu Leu Glu Gly Leu Leu Val  
35 40 45

Leu Glu His Cys Ala Thr Met Ala Trp Asp Cys Leu Met Arg Leu Glu  
50 55 60

55 Leu Leu Lys Arg Leu  
65

60 (2) INFORMATION FOR SEQ ID NO: 604:

701

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 604:

5 Lys Ile Val Tyr Ile Leu Gly Asn Pro Leu Lys Phe Asn Ser Arg Val  
1 5 10 15

10 Ile His His Leu Val Leu Leu Gln  
20

## (2) INFORMATION FOR SEQ ID NO: 605:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 605:

25 Met Asn Leu His Gln Arg Arg Leu Leu Ile Gly His Leu Met Thr  
1 5 10 15

30 Leu Val Lys Ala Ser Lys Ser Phe Ser Phe Thr Glu Ile Thr Ser Ser  
20 25 30

Arg Lys Lys  
35

## (2) INFORMATION FOR SEQ ID NO: 606:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 130 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 606:

40 Leu Leu Gly Tyr Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr  
1 5 10 15

45 Tyr Asn Ile His Leu His Ala Leu Phe Tyr Leu Phe Thr Leu Leu Val  
20 25 30

50 Gly Gly Leu Ser Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr  
35 40 45

Val Gly Pro Thr Gln Arg Leu Leu Cys Gly Thr Leu Ala Ala Leu  
50 55 60

55 His Met Leu Phe Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Xaa  
65 70 75 80

Glu Gly Ile Leu Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln  
85 90 95

60

702

Arg Val Pro Arg Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro  
100 105 110

5 Thr Thr Val Leu Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln  
115 120 125

Ser His  
130

10

## (2) INFORMATION FOR SEQ ID NO: 607:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 607:

20 Met Leu Val Ile Phe Leu Phe Thr Ser Leu Leu Lys Ile Pro Ser Ser  
1 5 10 15

25 Val Pro Gly Leu Ile Asn Val  
20

## (2) INFORMATION FOR SEQ ID NO: 608:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 608:

35 Glu Leu Asp Tyr Ile Leu  
1 5

## (2) INFORMATION FOR SEQ ID NO: 609:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 232 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 609:

50 Met Ala Pro Pro Gly Trp Gln Xaa Xaa Xaa Xaa Trp Leu Ala Cys  
1 5 10 15

Pro Asp Arg Gly Glu Leu Ser Ser Arg Ser Pro Pro Cys Arg Leu Ala  
20 25 30

55 Arg Trp Ala Glu Gly Asp Arg Glu Thr Arg Thr Cys Leu Leu Glu Leu  
35 40 45

Ser Ala Gln Ser Trp Gly Gly Arg Phe Arg Arg Ser Ser Ala Val Ser  
50 55 60

60

703

- Ala Gly Ser Pro Ser Arg Leu His Phe Leu Pro Gln Pro Leu Leu Leu  
65 70 75 80  
Arg Ser Ser Gly Ile Pro Ala Ala Thr Pro Trp Pro Gln Pro Ala  
85 90 95  
Gly Leu Pro Val Arg Pro Thr Thr Arg Thr Gly Gln Asp Arg  
100 105 110  
10 Thr Leu Asp Ile Ser Ile Cys Thr Gln Val Leu Ala Gly Thr Gln  
115 120 125  
Pro Pro Pro Arg Met Thr Ser Pro Ser Ser Pro Val Phe Arg  
130 135 140  
15 Leu Gln Thr Leu Asp Gly Gly Gln Gln Asp Gly Ser Gln Ala Asp Arg  
145 150 155 160  
Gly Lys Leu Asp Phe Gly Ser Gly Leu Pro Pro Met Gln Ser Gln Phe  
165 170 175  
20 Gln Gly Gln Asp Arg Lys Phe Ala Pro Ser Asp Lys Ser Gln Pro Pro  
180 185 190  
25 Thr Thr Gln Arg Gln Gln Val Pro Val Ser Arg Ile Gln Thr Asp Leu  
195 200 205  
Thr Gln Ile Gly Ser Ser Met Arg Ser Pro Gly Val Ser Pro Arg Ile  
210 215 220  
30 Trp Leu Asp Phe Gln Ser Thr Xaa  
225 230  
35 (2) INFORMATION FOR SEQ ID NO: 610:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 610:  
Met Val Leu Leu Leu Leu Ala Tyr Val Leu Leu Thr Tyr Ile Leu  
45 1 5 10 15  
Leu Leu Asn Met Leu Ile Ala Leu Met Xaa Arg Asp Arg Gln Gln Cys  
20 25 30  
50 Arg His
- 55 (2) INFORMATION FOR SEQ ID NO: 611:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- 60

704

- (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 611:  
Met Val Phe Gln Gly Phe Ser Ser Ala Phe Cys Leu Ser Ser Thr Ala  
1 5 10 15  
Pro Thr Ser His Pro  
20  
5  
10 (2) INFORMATION FOR SEQ ID NO: 612:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 612:  
Gly Lys Lys Asn Gln Leu Leu Val Ile  
1 5  
20  
25 (2) INFORMATION FOR SEQ ID NO: 613:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 613:  
Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val  
1 5 10 15  
30 Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys  
20 25  
35 (2) INFORMATION FOR SEQ ID NO: 614:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 614:  
Met Ala Lys Arg Ser Pro Gly Gly Cys Gly Ser Gly Leu Ile Leu Leu  
45 1 5 10 15  
Cys Cys Gln Pro Cys Arg Pro Thr Ser Ser Ala Pro Met Arg  
20 25 30  
50  
55 (2) INFORMATION FOR SEQ ID NO: 615:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 113 amino acids  
(B) TYPE: amino acid
- 60



## (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 615:

5 Ile Thr Ile Ala Ile Gln Met Ile Cys Leu Val Asn Xaa Glu Leu Tyr 15  
1 5 10  
Pro Thr Phe Val Arg Asn Xaa Gly Val Met Val Cys Ser Ser Leu Cys 30  
20 25  
10 Asp Ile Gly Gly Ile Ile Thr Pro Phe Ile Val Phe Arg Leu Arg Glu 45  
35 40  
Val Trp Gln Ala Leu Pro Leu Ile Leu Phe Ala Val Leu Gly Leu Leu 60  
50 55  
15 Ala Ala Gly Val Thr Leu Leu Leu Pro Glu Thr Lys Gly Val Ala Leu 80  
65 70 75  
Pro Glu Thr Met Lys Asp Ala Glu Asn Leu Leu Gly Arg Lys Ala Lys Pro 95  
85 90  
20 Lys Glu Asn Thr Ile Tyr Leu Lys Val Gln Thr Ser Glu Pro Ser Gly 110  
100 105

25 Thr

30 (2) INFORMATION FOR SEQ ID NO: 616:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 616:

35 Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro Lys Glu 15  
1 5 10  
40 Asn Thr

45 (2) INFORMATION FOR SEQ ID NO: 617:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 617:

55 Pro Arg Val Arg Asn Ser Pro Glu Asp Leu Gly Leu Ser Leu Thr Gly 15  
1 5 10  
Asp Ser Cys Lys Leu 20

60

(2) INFORMATION FOR SEQ ID NO: 618:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 618:

5 Gln Ala Asp Asp Leu Gln Ala Thr Val Ala Ala Leu Cys Val Leu Arg 15  
1 5 10  
10 Gly Gly Gly Pro Trp Ala Gly Ser Trp Leu Ser Pro Lys Thr Pro Gly 30  
20 25  
15 Ala Met Gly Gly Asp Leu Val Leu Gly Ala Leu Arg Arg Arg 45  
35 40  
20 Lys Arg Leu Leu 50

25 (2) INFORMATION FOR SEQ ID NO: 619:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 232 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 619:

30 Glu Gln Glu Lys Ser Leu Ala Gly Trp Ala Leu Val Leu Ala Xaa 15  
1 5 10  
35 Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp Phe Gly Gly 30  
20 25  
Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr Leu Trp Leu 45  
35 40  
40 Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val Val Pro Gly 60  
50 55  
45 Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val Met Gly Val 80  
65 70 75  
Cys Cys Thr Ala Leu Leu Val Ala Val Ala Arg Lys Leu Glu Phe 95  
85 90  
50 Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp Ile Gln Tyr 110  
100 105  
Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln Glu Ala Trp 125  
115 120  
55 Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala Ala Arg Xaa 140  
130 135  
His Gln Arg Xaa Leu Leu Ala Ala Ile Asn Ala Phe Arg Gln Val Arg 160  
145 150 155

707

- Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met Val Asp Ile  
165 170 175
- 5 Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn Leu Ser Ser  
180 185 190
- Ser His Arg Ala Leu Glu Lys Lys Gln Ile Asp Thr Leu Ala Gly Lys Leu  
195 200 205
- 10 Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly Pro Arg Gln Leu  
210 215 220
- 15 Pro Glu Pro Ser Gln Gln Ser Lys  
225 230
- 20 (2) INFORMATION FOR SEQ ID NO: 620:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 620:
- 25 Tyr Gln Ala His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly  
1 5 10  
Thr Arg Gly Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser  
20 25 30
- 35 Gly Ala Gly Lys  
35
- 40 (2) INFORMATION FOR SEQ ID NO: 621:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 621:
- 45 Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu  
1 5 10  
Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
20 25 30
- 50 Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala  
35 40 45
- 55 His Lys Ala Lys Ser His Pro Glu Val  
50 55
- 60 (2) INFORMATION FOR SEQ ID NO: 622:

708

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 622:
- 5 Ile Thr Ser Thr Asp Ile Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln  
1 5 10  
Pro Ser Asp
- 10 (2) INFORMATION FOR SEQ ID NO: 623:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 623:
- 15 Asn Ser Thr Ser Gly Glu Cys Leu Leu Leu Glu Ala Glu Gly Met Ser  
1 5 10  
Lys Ser Tyr
- 20 (2) INFORMATION FOR SEQ ID NO: 624:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 51 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 624:
- 25 Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys Ile Phe  
1 5 10  
Val Gly Ser Gly Ser Ser Gly Gly Thr Glu Gly Leu Val Met Asn Ser  
20 25 30
- 30 Asp Ile Leu Gly Ala Thr Thr Glu Val Leu Ile Glu Asp Ser Asp Ser  
35 40 45
- 35 Ala Gly Pro  
50
- 40 (2) INFORMATION FOR SEQ ID NO: 625:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 60 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 625:
- 45
- 50
- 55
- 60

WO 98/39448

710

Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala  
 35 40 45  
 5 His Lys Ala Lys Ser His Pro Glu Val Xaa Ile Thr Ser Thr Asp Ile  
 50 55 60  
 Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln Pro Ser Asp Xaa Asn Ser  
 65 70 75 80  
 10 Thr Ser Gly Glu Cys Leu Leu Glu Ala Glu Gly Met Ser Lys Ser  
 85 90 95  
 Tyr Xaa Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys  
 100 105 110  
 Ile Phe Val Gly Ser Gly Ser Ser Gly Thr Glu Gly Leu Val Met  
 115 120 125  
 20 Asn Ser Asp Ile Leu Gly Ala Thr Thr Glu Val Leu Ile Glu Asp Ser  
 130 135 140  
 Asp Ser Ala Gly Pro Xaa Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala  
 145 150 155 160  
 25 Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met Ile His  
 165 170 175  
 Thr Gly Glu Lys His Tyr Xaa  
 180

(2) INFORMATION FOR SEQ ID NO: 629:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 60 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 629:

Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His  
 1 5 10 15  
 45 Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Lys Lys  
 20 25 30  
 Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln  
 35 40 45  
 Lys Gln Leu Leu Arg His Ala Lys His His Thr Asp  
 50 55 60

(2) INFORMATION FOR SEQ ID NO: 630:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 amino acids
  - (B) TYPE: amino acid

PCT/US98/04493

WO 98/39448

709

Ile Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala  
 1 5 10 15  
 5 His Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Leu Lys  
 20 25 30  
 Lys Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu  
 35 40 45  
 10 Gln Lys Gln Leu Leu Arg His Ala Lys His His Thr  
 50 55 60  
 15 (2) INFORMATION FOR SEQ ID NO: 626:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 626:  
 Asp Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala Arg Ala Phe Lys Ser  
 1 5 10 15  
 Ser His Asn Leu Ala Val His Arg Met Ile His Thr Gly Glu Lys  
 20 25 30  
 30

(2) INFORMATION FOR SEQ ID NO: 627:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 627:

Arg Ser Ser Arg Ser Lys Thr Gly Ser Leu Gln Leu Ile Cys Lys Ser  
 1 5 10 15  
 Glu Pro Asn Thr Asp Gln Leu Asp Tyr  
 20 25

(2) INFORMATION FOR SEQ ID NO: 628:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 183 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 628:

Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu  
 1 5 10 15  
 Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
 20 25 30

711

(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 630:

5 Pro Phe Lys Asp Asp Pro Arg Asp Glu Thr Tyr Lys Pro His Leu Glu  
1 5 10 15  
Arg Glu Thr Pro Lys Pro Arg Arg Lys Ser Gly  
20 25

10 (2) INFORMATION FOR SEQ ID NO: 631:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 110 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 631:

20 Glu Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Leu  
1 5 10 15  
Met Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val  
20 25 30  
Leu Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu  
35 40 45  
Trp Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg  
50 55 60  
Met Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val  
65 70 75 80  
Gln Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu  
85 90 95  
Asn Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met  
100 105 110

40 (2) INFORMATION FOR SEQ ID NO: 632:

45 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 632:

50 Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg Pro Thr Gly Gln Gln  
1 5 10 15  
Leu Glu Ser Leu Gly Leu Leu Ala  
20

60 (2) INFORMATION FOR SEQ ID NO: 633:

712

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 633:  
Val His Arg Glu Glu Ala Ser Cys Tyr Cys Gln Ala Glu Pro Ser Gly  
1 5 10 15  
Asp Leu  
20

15 (2) INFORMATION FOR SEQ ID NO: 634:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 634:

20 Arg Pro Ala Leu Arg Gln Ala Gly Gly Thr Arg Glu Pro Arg Gln  
1 5 10 15  
Lys Arg Trp Ala Gly Leu  
20

30 (2) INFORMATION FOR SEQ ID NO: 635:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 635:

40 Ala Val Asn Phe Arg Pro Gln Arg Ser Gln Ser Met  
1 5 10

45 (2) INFORMATION FOR SEQ ID NO: 636:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 636:

50 Met Ile Thr Asp Val Gln Leu Ala Ile Phe Ala Asn Met Leu Gly Val  
1 5 10 15  
Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val Ala Val Asn Asn  
20 25 30  
Pro Lys Lys Gln Glu  
35

60

## (2) INFORMATION FOR SEQ ID NO: 637:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 342 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 637:

10 Glu Glu Met Ala Asp Ser Val Lys Thr Phe Leu Gln Asp Leu Ala Arg  
 1 5 10 15

15 Gly Ile Lys Asp Ser Ile Trp Gly Ile Cys Thr Ile Ser Lys Leu Asp  
 20 25 30

Ala Arg Ile Gln Gln Lys Arg Glu Gln Arg Arg Arg Ala Ser  
 35 40 45

20 Ser Val Leu Ala Gln Arg Arg Ala Gln Ser Ile Glu Arg Lys Gln Glu  
 50 55 60

Ser Glu Pro Arg Ile Val Ser Arg Ile Phe Gln Cys Cys Ala Trp Asn  
 65 70 75 80

25 Gly Gly Val Phe Trp Phe Ser Leu Leu Phe Tyr Arg Val Phe Ile  
 85 90 95

Pro Val Leu Gln Ser Val Thr Ala Arg Ile Ile Gly Asp Pro Ser Leu  
 100 105 110

His Gly Asp Val Trp Ser Trp Leu Glu Phe Phe Leu Thr Ser Ile Phe  
 115 120 125

35 Ser Ala Leu Trp Val Leu Pro Leu Phe Val Leu Ser Lys Val Val Asn  
 130 135 140

Ala Ile Trp Phe Gln Asp Ile Ala Asp Leu Ala Phe Glu Val Ser Gly  
 145 150 155 160

Arg Lys Pro His Pro Phe Pro Ser Val Ser Lys Ile Ile Ala Asp Met  
 165 170 175

Leu Phe Asn Leu Leu Leu Gln Ala Leu Phe Leu Ile Gln Gly Met Phe  
 180 185 190

Val Ser Leu Phe Pro Ile His Leu Val Gly Gln Leu Val Ser Leu Leu  
 195 200 205

50 His Met Ser Leu Leu Tyr Ser Leu Tyr Cys Phe Glu Tyr Arg Trp Phe  
 210 215 220

Asn Lys Gly Ile Glu Met His Gln Arg Leu Ser Asn Ile Glu Arg Asn  
 225 230 235 240

Trp Pro Tyr Tyr Phe Gly Phe Gly Leu Pro Leu Ala Phe Leu Thr Ala  
 245 250 255

Met Gln Ser Ser Tyr Ile Ile Ser Gly Cys Leu Phe Ser Ile Leu Phe  
 260 265 270

Pro Leu Phe Ile Ile Ser Ala Asn Glu Ala Lys Thr Pro Gly Lys Ala  
 275 280 285

5 Tyr Leu Phe Gln Leu Arg Leu Phe Ser Leu Val Val Phe Leu Ser Asn  
 290 295 300

Arg Leu Phe His Lys Thr Val Tyr Leu Gln Ser Ala Leu Ser Ser Ser  
 305 310 315 320

10 Thr Ser Ala Glu Lys Phe Pro Ser Pro His Pro Ser Pro Ala Lys Leu  
 325 330 335

Lys Ala Thr Ala Gly His  
 340

(2) INFORMATION FOR SEQ ID NO: 638:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 529 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 638:

Met Ala Lys Phe Met Thr Pro Val Ile Gln Asp Asn Pro Ser Gly Trp  
 1 5 10 15

30 Gly Pro Cys Ala Val Pro Glu Gln Phe Arg Asp Met Pro Tyr Gln Pro  
 20 25 30

Phe Ser Lys Gly Asp Arg Leu Gly Lys Val Ala Asp Trp Thr Gly Ala  
 35 40 45

Thr Tyr Gln Asp Lys Arg Tyr Thr Asn Lys Tyr Ser Ser Gln Phe Gly  
 50 55 60

Gly Gly Ser Gln Tyr Ala Tyr Phe His Glu Glu Asp Glu Ser Ser Phe  
 65 70 75 80

Gln Leu Val Asp Thr Ala Arg Thr Gln Lys Thr Ala Tyr Gln Arg Asn  
 85 90 95

45 Arg Met Arg Phe Ala Gln Arg Asn Leu Arg Arg Asp Lys Asp Arg Arg  
 100 105 110

Asn Met Leu Gln Phe Asn Leu Gln Ile Leu Pro Lys Ser Ala Lys Gln  
 115 120 125

Lys Glu Arg Glu Arg Ile Arg Leu Gln Lys Lys Phe Gln Lys Gln Phe  
 130 135 140

Gly Val Arg Gln Lys Trp Asp Gln Lys Ser Gln Lys Pro Arg Asp Ser  
 145 150 155 160

Ser Val Glu Val Arg Ser Asp Trp Glu Val Lys Glu Met Asp Phe  
 165 170 175

60 Pro Gln Leu Met Lys Met Arg Tyr Leu Glu Val Ser Glu Pro Gln Asp

180 185 190  
Ile Glu Cys Cys Gly Ala Leu Glu Tyr Tyr Asp Lys Ala Phe Asp Arg  
195 200 205  
Ile Thr Thr Arg Ser Glu Lys Pro Leu Arg Xaa Xaa Lys Arg Ile Phe  
210 215 220  
His Thr Val Thr Thr Thr Asp Asp Pro Val Ile Arg Lys Leu Ala Lys  
225 230 235 240  
Thr Glu Gly Asn Val Phe Ala Thr Asp Ala Ile Leu Ala Thr Leu Met  
245 250 255  
15 Ser Cys Thr Arg Ser Val Tyr Ser Trp Asp Ile Val Val Glu Arg Val  
260 265 270  
Gly Ser Lys Leu Phe Phe Asp Lys Arg Asp Asn Ser Asp Phe Asp Leu  
275 280 285  
Leu Thr Val Ser Glu Thr Ala Asn Glu Pro Pro Glu Asp Glu Gly Asn  
290 295 300  
25 Ser Phe Asn Ser Pro Arg Asn Leu Ala Met Glu Ala Thr Tyr Ile Asn  
305 310 315 320  
His Asn Phe Ser Glu Glu Cys Leu Arg Met Gly Lys Glu Arg Tyr Asn  
325 330 335  
30 Phe Pro Asn Pro Asn Pro Phe Val Glu Asp Asp Met Asp Lys Asn Glu  
340 345 350  
Ile Ala Ser Val Ala Tyr Arg Tyr Arg Ser Gly Lys Leu Gly Asp Asp  
355 360 365  
35 Ile Asp Leu Ile Val Arg Cys Glu His Asp Gly Val Met Thr Gly Ala  
370 375 380  
Asn Gly Glu Val Ser Phe Ile Asn Ile Lys Thr Leu Asn Glu Trp Asp  
385 390 395 400  
Ser Arg His Cys Asn Gly Val Asp Trp Arg Glu Lys Leu Asp Ser Glu  
405 410 415  
45 Arg Gly Ala Val Ile Ala Thr Glu Leu Lys Asn Asn Ser Tyr Lys Leu  
420 425 430  
Ala Arg Trp Thr Cys Cys Ala Leu Leu Ala Gly Ser Glu Tyr Leu Lys  
435 440 445  
50 Leu Gly Tyr Val Ser Arg Tyr His Val Lys Asp Ser Ser Arg His Val  
450 455 460  
Ile Leu Gly Thr Glu Glu Phe Lys Pro Asn Glu Phe Ala Ser Glu Ile  
465 470 475 480  
Asn Leu Ser Val Glu Asn Ala Trp Gly Ile Leu Arg Cys Val Ile Asp  
485 490 495  
60 Ile Cys Met Lys Leu Glu Glu Gly Lys Leu Ile Leu Lys Asp Pro

500 505 510  
Asn Lys Glu Val Ile Arg Val Tyr Ser Leu Pro Asp Gly Thr Phe Ser  
515 520 525  
5 Ser  
505  
10 (2) INFORMATION FOR SEQ ID NO: 639:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 194 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 639:  
20 Lys Lys Arg His Thr Asp Val Glu Glu Phe Tyr Thr Glu Val Gly Glu Ile  
1 5 10 15  
Thr Thr Asp Leu Gly Lys His Glu His Met His Asp Arg Asp Asp Leu  
20 25 30  
25 Tyr Ala Glu Glu Met Glu Arg Glu Met Arg His Lys Leu Lys Thr Ala  
35 40 45  
Phe Lys Asn Phe Ile Glu Lys Val Glu Ala Leu Thr Lys Glu Glu Leu  
50 55 60  
30 Glu Phe Glu Val Pro Phe Arg Asp Leu Gly Phe Asn Gly Ala Pro Tyr  
65 70 75 80  
Arg Ser Thr Cys Leu Leu Glu Pro Thr Ser Ser Ala Leu Val Asn Ala  
85 90 95  
35 Thr Glu Trp Pro Pro Phe Val Val Thr Leu Asp Glu Val Glu Leu Ile  
100 105 110  
40 His Phe Xaa Arg Val Glu Phe His Leu Lys Asn Phe Asp Met Val Ile  
115 120 125  
Val Tyr Lys Asp Tyr Ser Lys Lys Val Thr Met Ile Asn Ala Ile Pro  
130 135 140  
45 Val Ala Ser Leu Asp Pro Ile Lys Glu Trp Leu Asn Ser Cys Asp Leu  
145 150 155 160  
Lys Tyr Thr Glu Gly Val Glu Ser Ser Leu Asn Trp Thr Lys Ile Met Lys  
165 170 175  
Thr Ile Val Asp Asp Pro Glu Gly Phe Phe Glu Glu Gly Trp Ser  
180 185 190  
55 Phe Leu  
60 (2) INFORMATION FOR SEQ ID NO: 640:

717

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 640:

Arg Ser Gly Leu Gly Leu Gly Ile Thr Ile Ala Phe Leu Ala Thr Leu  
1 5 10 15

Ile Thr Gln Phe Leu Val Tyr Asn Gly Val Tyr Gln Tyr Thr Ser Pro  
20 25 30

Asp Phe Leu Tyr Ile Arg Ser Trp Leu Pro Cys Ile Phe Phe Ser Gly  
35 40 45

Gly Val Thr Val Gly Asn Ile Gly Arg Gln Leu Ala Met Gly Val Pro  
50 55 60

Glu Lys Pro His Ser Asp  
65 70

## (2) INFORMATION FOR SEQ ID NO: 641:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 101 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 641:

Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser Val Glu  
1 5 10 15

Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Xaa Xaa Pro Xaa  
20 25 30

Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser Phe Tyr  
35 40 45

Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg Leu Phe  
50 55 60

Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile Ser Asn  
65 70 75 80

Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu Leu Asp  
85 90 95

Thr Arg Arg Ser Gly  
100

## (2) INFORMATION FOR SEQ ID NO: 642:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 233 amino acids

(B) TYPE: amino acid

718

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 642:

Met Glu Ala Gln Gln Val Asn Glu Ala Glu Ser Ala Arg Glu Gln Leu  
1 5 10 15

Gln Xaa Leu His Asp Gln Ile Ala Gly Gln Lys Ala Ser Lys Gln Glu  
20 25 30

Leu Glu Thr Glu Leu Glu Arg Leu Lys Gln Glu Phe His Tyr Ile Glu  
35 40 45

Glu Asp Leu Tyr Arg Thr Lys Asn Thr Leu Gln Ser Arg Ile Lys Asp  
50 55 60

Arg Asp Glu Glu Ile Gln Lys Leu Arg Asn Gln Leu Thr Asn Lys Thr  
65 70 75 80

Leu Ser Asn Ser Ser Gln Ser Glu Leu Glu Asn Arg Leu His Gln Leu  
85 90 95

Thr Glu Thr Leu Ile Gln Lys Gln Thr Met Leu Glu Ser Leu Ser Thr  
100 105 110

Glu Lys Asn Ser Leu Val Phe Gln Leu Glu Arg Leu Glu Gln Met  
115 120 125

Asn Ser Ala Ser Gly Ser Ser Ser Asn Gly Ser Ser Ile Asn Met Ser  
130 135 140

Gly Ile Asp Asn Gly Glu Gly Thr Arg Leu Arg Asn Val Pro Val Leu  
145 150 155 160

Phe Asn Asp Thr Glu Thr Asn Leu Ala Gly Met Tyr Gly Lys Val Arg  
165 170 175

Lys Ala Ala Ser Ser Ile Asp Gln Phe Ser Ile Arg Leu Gly Ile Phe  
180 185 190

Leu Arg Arg Tyr Pro Ile Ala Arg Val Phe Val Ile Ile Tyr Met Ala  
195 200 205

Leu Leu His Leu Trp Val Met Ile Val Leu Leu Thr Tyr Thr Pro Glu  
210 215 220

Met His His Asp Gln Pro Tyr Gly Lys  
225 230

50

(2) INFORMATION FOR SEQ ID NO: 643:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 643:

Ile Arg His Glu Gln His Pro Asn Phe Ser Leu Glu Met His Ser Lys  
1 5 10 15

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97897	
C. ADDITIONAL INDICATIONS (name blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (name blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only

☒ This sheet was received with the international application

For International Bureau use only

☐ This sheet was received by the International Bureau on:

Authorized officer

Susan White

PCT International Division

Authorized officer



## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209043	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

<input checked="" type="checkbox"/> This sheet was received with the international application	For receiving Office use only
Authorized officer Susan White PCT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 119, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit September 4, 1997 Accession Number 209235	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

<input checked="" type="checkbox"/> This sheet was received with the international application	For receiving Office use only
Authorized officer Susan White PCT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer

WO 98/39448

723

PCT/US98/04493

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97898
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

WO 98/39448

724

PCT/US98/04493

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209044
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b1a)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97899	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b1a)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209045	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>130</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit April 28, 1997	Accession Number 209011
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the International application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>131</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97900
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the International application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97901	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page 131, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209046	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application		<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer Susan White PCT International Division		Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136b)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209047
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136b)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 22, 1997	Accession Number 209076
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b(ii))

A. The indications made below relate to the microorganism referred to in the description on page 140, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit August 21, 1997 Accession Number 209215 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer
--	---

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13b(ii))

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depositary institution American Type Culture Collection Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit February 26, 1997 Accession Number 97904 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer
--	---

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page <u>154</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet <input type="checkbox"/>
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit July 3, 1997	Accession Number 209139
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
<div style="display: flex; justify-content: space-between;"> <div> <p>For receiving Office use only</p> <p><input checked="" type="checkbox"/> This sheet was received with the international application</p> <p>Authorized officer Susan White PCT International Division</p> </div> <div> <p>For International Bureau use only</p> <p><input type="checkbox"/> This sheet was received by the International Bureau on:</p> <p>Authorized officer</p> </div> </div>	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page <u>153</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet <input type="checkbox"/>
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209049
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
<div style="display: flex; justify-content: space-between;"> <div> <p>For receiving Office use only</p> <p><input checked="" type="checkbox"/> This sheet was received with the international application</p> <p>Authorized officer Susan White PCT International Division</p> </div> <div> <p>For International Bureau use only</p> <p><input type="checkbox"/> This sheet was received by the International Bureau on:</p> <p>Authorized officer</p> </div> </div>	



## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 133, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution: American Type Culture Collection Address of depository institution (including postal code and country): 12301 Parklawn Drive, Rockville, Maryland 20852, United States of America Date of deposit: February 26, 1997 Accession Number: 97903	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer: Susan White, PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer:	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution: American Type Culture Collection Address of depository institution (including postal code and country): 12301 Parklawn Drive, Rockville, Maryland 20852, United States of America Date of deposit: June 12, 1997 Accession Number: 209119	
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer: Susan White, PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer:	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97902
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
For International Bureau use only	
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209048
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
For International Bureau use only	
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 15, 1997
Accession Number	209050
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:		
Authorized officer	Susan White PCT International Division	Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 12, 1998
Accession Number	209627
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:		
Authorized officer	Susan White PCT International Division	Authorized officer	

*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
    - (f) a polynucleotide which is a variant of SEQ ID NO:X;
    - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
    - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
    - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
  2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
  3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
  - (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z.

745

(g) a variant of SEQ ID NO: Y;

(h) an allelic variant of SEQ ID NO: Y; or

(i) a species homologue of the SEQ ID NO: Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

5

13. A method of identifying a binding partner to the polypeptide of claim 11 comprising:

10

14.

expresses the isolated polypeptide of claim 11 in a cell;

11.

15. A method of identifying a binding partner to the polypeptide of claim 11 comprising:

(a) culturing the recombinant cell;

(b) recovering said polypeptide.

under conditions such that the polypeptide is expressed; and

(c) identifying the protein in the supernatant having the activity.

comprising:

(a) expressing SEQ ID NO: X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

16. The polypeptide produced by the method of claim 22.

20

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

25

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

30

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

35

746

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

5

21. The gene corresponding to the cDNA sequence of SEQ ID NO: Y.

21. The gene corresponding to the cDNA sequence of SEQ ID NO: Y.

10

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO: X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

15

23. The product produced by the method of claim 22.